

On Marginal Quasi-Likelihood Inference in Generalized Linear Mixed Models

Brajendra C. Sutradhar¹

Memorial University of Newfoundland, St. John's, Canada

and

R. Prabhakar Rao

Sri Sathya Sai Institute of Higher Learning, India

View metadata, citation and similar papers at core.ac.uk

In view of the cumbersome and often intractable numerical integrations required for a full likelihood analysis, several suggestions have been made recently for approximate inference in generalized linear mixed models (GLMMs). Two closely related approximate methods are the penalized quasi-likelihood (PQL) method and the marginal quasi-likelihood (MQL) method. The PQL approach generally produces biased estimates for the regression effects and the variance component of the random effects. Recently, some corrections have been proposed to remove these biases. But the corrections appear to be satisfactory only when the variance component of the random effects is small. The MQL approach has also been used for inference in the GLMMs. This approach requires the computations of the joint moments of the clustered observations, up to order four. But the derivation of these moments are not easy. Consequently, different “working” formulas have been used, especially for the mean and covariance matrix of the observations, which may not lead to desirable estimates. In this paper, we use a small variance component (of the random effects) approach and develop the MQL estimating equations for the parameters based on the joint moments of order up to four. The proposed approach thus avoids the use of the so-called “working” covariance and higher order moment matrices, leading to better estimates for the regression and the over-dispersion parameters, in the sense of efficiency in particular. © 2000 Academic Press

AMS 1991 subject classifications: Primary 62A10; Secondary 62F11, 62F12.

Key words and phrases: regression parameters, variance component of the random effects, quasi-likelihood estimation, consistency, improved efficiency.

¹ To whom correspondence should be addressed: B. C. Sutradhar, Department of Mathematics and Statistics, Memorial University of Newfoundland, St. John's, NF, Canada A1C 5S7.

1. INTRODUCTION

Generalized linear mixed models (GLMMs) are useful for accommodating the overdispersion and correlations often observed among outcomes. These models are generated from the well-known generalized linear model (GLM) (McCullagh and Nelder, 1989) by adding random effects to the linear predictor. Unfortunately, a full likelihood analysis in GLMMs is often hampered by the need for numerical integration. To overcome such integration problems, several suggestions have been made recently for approximate inference in GLMMs. Breslow and Clayton (1993), for example, consider two closely related approximate methods, namely the penalized quasi-likelihood (PQL) method and the marginal quasi-likelihood (MQL) method for inferences in GLMMs. The PQL approach of Breslow and Clayton (1993) generally produces biased estimates for the regression effects and variance component of the random effects, the biases in variance component estimates being considerably larger as compared to the biases in regression estimates. For a single source of extraneous variation, Breslow and Lin (1995) provided a correction factor for the estimates of this variance component derived from the Laplace approximations (Solomon and Cox, 1992) as well as PQL. They also provide a first-order correction term for the regression coefficients estimated by PQL (see also Goldstein and Rasbash, 1996, for similar but different improvements). The bias-corrected PQL estimators due to Breslow and Lin (1995) appear to improve the asymptotic performance of the uncorrected quantities. But the improvement is satisfactory only when the true variance component is small, more specifically, less than or equal to 0.25. Note that although it is generally recognized (cf. Sutradhar and Das, 1995) that the mixed model with small variance components of the random effects is a reasonable model, there are, however, situations where the variance component may be larger than 0.25. For example, consider an overdispersed Poisson model with mean $\mu = 1$, say. In this case the variance of the response will be $\mu + \sigma^2\mu^2$, where σ^2 is the variance of the random effects involved in the model. Under this mixed Poisson model there is no reason why the overdispersion cannot be greater than 0.25. Thus, in some practical situations, the variance component may be larger than 0.25, although the range $0 \leq \sigma^2 \leq 0.25$ is reasonably wide (Breslow and Lin, 1995, p. 90).

Following the generalized estimating equation approach of Zeger *et al.* (1988), Breslow and Clayton (1993), as mentioned above, also use the MQL method in estimating the regression effects of the GLMMs. The application of the estimating equation approach for the regression parameters requires the first and second order marginal moments of the responses. The exact first and second order moments of the responses under the GLMMs are, however, not available. Breslow and Clayton

(1993) have used an approximate mean vector and a “working covariance” matrix as in Zeger *et al.* (1988) to construct the estimating equations for the regression parameters (see also Waclawiw and Liang, 1993). Similar to the PQL approach, this “working” covariance based MQL approach also produces biased estimates for the regression effects (see also Rodriguez and Goldman, 1995). But, as shown by Sutradhar and Qu (1998), the amount of bias (produced by the MQL approach of Waclawiw and Liang, 1993, in particular) is considerably large only for the regression intercept estimator. With regard to the standard errors of the MQL estimators, they are, in general, larger than the corresponding PQL estimators for all regression parameters including the intercept.

The estimation of the variance component of the random effects in the GLMMs is, however, much more complex as compared to the regression estimation. This is because the construction of the estimating equations for the variance parameter requires the moments of the responses up to order four. Breslow and Clayton (1993), consequently, have used an alternative pseudolikelihood method due to Carroll and Ruppert (1982). More specifically, under the assumption that the mean vector of the responses is known, these authors use the normal theory likelihood based on a first-order variance approximation to estimate the variance component. The performance of the variance estimator in this approach may not be satisfactory. Also, this approach may produce a negative estimate of the variance component, which is undesirable (cf. Zeger, 1988, Section 3.2; Breslow and Clayton, 1993, Sections 3.3 and 4.1, Sutradhar and Qu, 1998, pp. 172 and 176). Such behavior of the variance component estimator would be anticipated from all the casual approximations used in this approach, such as approximation to the mean, using the approximate working covariance matrix and the normal approximation to an unknown distribution.

In this article, we provide a comprehensive theoretical treatment of the problem of MQL estimations for both the regression and variance parameters of the GLMMs. Similar to Breslow and Lin (1995), we consider a single source of extraneous variation. Given an unobserved random effect, observations in a cluster are assumed to be conditionally independent with exponential family form marginal density (cf. Liang and Zeger, 1986; Waclawiw and Liang, 1993; and Breslow and Lin, 1995). The random effects are assumed to be independently and identically distributed so that the fifth and the other higher order moments are negligible. For example, if the random effects are assumed to have normal distribution with zero mean and variance σ^2 (say), then the above assumption about the moments will accommodate relatively larger values of σ^2 such that for all $r \geq 5$ the r th moment of the random effects will be of order σ^5 . Consequently, this will provide a higher order likelihood approximation for the observations in a cluster, as compared to Breslow and Lin (1995), for example. Our technique of the

derivation for the likelihood approximation is different than that of Breslow and Lin (1995). The approximate likelihood function is derived by averaging over the distribution of the random effects (cf. Sutradhar and Das, 1995; Das and Sutradhar, 1996; Sutradhar and Rao, 1996; Lin, 1997).

The MQL estimating equations for the regression parameters and the variance component of the GLMMs are constructed in Sections 3 and 4 respectively. The construction of the MQL estimating equations requires moments of the responses up to order four to be known. By direct exploitation of the joint density, and after some lengthy algebras, we provide these moments up to order two in Section 3 and the remaining third and the fourth order moments in Section 4. Note that the formulas for the moments up to order four for the GLMMs are interesting on their own right. Further, note that in the present approach it is not only that the MQL method uses the almost exact means of the responses but it also completely avoids the use of the working covariance, yielding consistent as well as more efficient estimates of the parameters involved. To shed some light on the efficiency improvement due to the proposed MQL approach we also provide, in Section 3, a numerical comparison between the variances of the regression estimates obtained with the existing and the proposed MQL approaches. With respect to the estimation of the variance component of the GLMMs, the present MQL approach, unlike the existing approaches (cf. Prentice and Zhao, 1991), uses the almost exact second, third, and the fourth order moments of the responses. In Section 5, we consider the testing for the significance of the variance component of the random effects. If the null hypothesis that “variance of the random effects is zero” is accepted, then the GLMMs would reduce to the generalized linear fixed effects model. We discuss both Wald and the score tests for testing such hypotheses. Further, in Section 6 we consider the testing for the homogeneity of the variance components of several conceptual groups, where the clustered observations in each group follow the GLMMs with the same regression parameters but possibly different variance components. Lin (1997) considers a similar but different testing problem. The conclusion of the paper is given in Section 7.

2. HIGHER ORDER LIKELIHOOD APPROXIMATION

Consider a clustered data set consisting of a response y_{ij} for the j th ($j = 1, \dots, n_i$) individual on the i th ($i = 1, \dots, I$) cluster and a $p \times 1$ vector x_{ij} of covariates associated with that response. Let β denote a $p \times 1$ vector of unknown fixed effect parameters associated with the covariate x_{ij} . Further, let γ_i be a univariate random effect such that for a given γ_i , n_i observations in the i th cluster are independent and they have the conditional density

$$\begin{aligned}
L_i^*(\beta, \gamma_i) &= \Pi_{j=1}^{n_i} f_{ij}^*(y_{ij} | \theta_{ij}^*) \\
&= \Pi_{j=1}^{n_i} \exp[\{y_{ij}\theta_{ij}^* - a(\theta_{ij}^*)\}\phi + c(y_{ij}, \phi)] \quad (2.1)
\end{aligned}$$

(cf. Breslow and Lin, 1995, p. 82), where $f_{ij}^*(\cdot)$ is of the exponential form, $\theta_{ij}^* = \theta_{ij} + \{b_i(\theta_{ij})\}^{1/2} \gamma_i$, with $\theta_{ij} = x_{ij}^T \beta$, and $b_i(\cdot)$ is a suitable bounded function, $a(\cdot)$ and $c(\cdot)$ are known functional forms, and ϕ is a possibly unknown scale parameter. We do not make any specific distributional assumption for the random effects γ_i ($i = 1, \dots, I$). It is rather assumed that

$$E\gamma_i^r = \delta_r(\sigma^2) = \sum_{s=1}^r c_{r,s} \sigma^{r+1-s}, \quad \text{for } r = 1, \dots, 4, \quad (2.2)$$

and

$$E\gamma_i^r = o(\sigma^r), \quad \text{for } r \geq 5,$$

where $c_{r,s}$ are suitable known constants for $r = 1, \dots, 4$ with $c_{1,1} = c_{2,2} = 0$, and $c_{2,1} = 1$. For the case when $\gamma_i \sim N(0, \sigma^2)$, $c_{1,1}$, $c_{2,1}$, and $c_{2,2}$ satisfy the stated conditions, and the remaining $c_{r,s}$ are zero; that is, $c_{3,1} = c_{3,2} = c_{3,3} = 0$ and $c_{4,1} = 3$, $c_{4,2} = c_{4,3} = c_{4,4} = 0$.

Under the assumption that $\gamma_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$, Breslow and Lin (1995) approximated the likelihood for the observed data

$$\begin{aligned}
L(\beta, \sigma^2) &= \Pi_{i=1}^I L_i(\beta, \sigma^2) \\
&= \Pi_{i=1}^I (2\pi\sigma^2)^{-1/2} \int \exp[\log L_i^*(\beta, \gamma_i) - \gamma_i^2/2\sigma^2] d\gamma_i, \quad (2.3)
\end{aligned}$$

first by making a quartic expansion of the integrand in (2.3) in a Taylor series about its maximum value $\tilde{\gamma}_i$ (say), and then by taking appropriate expectations involved in the series expansion. The resulting approximate likelihood was maximized to obtain the likelihood estimates of β and σ^2 .

Based on the assumption made in (2.2), we now obtain the unconditional joint density for the data. More specifically, we expand the conditional density $f_{ij}^*(y_{ij} | \theta_{ij}^*)$ in (2.1) about θ_{ij} and take the expectation over θ_{ij}^* under the assumption that $E(\gamma_i^r) = o(\sigma^r)$, for $r \geq 5$. This, for $c_{1,1} = 0$, $c_{2,1} = 1$, and $c_{2,2} = 0$, yields the approximate likelihood for the data as

$$\begin{aligned}
L_i(\beta, \sigma^2) &= [\Pi_{j=1}^{n_i} f_{ij}(y_{ij} | \theta_{ij})] \\
&\times \left[1 + \phi \left[\frac{\sigma^2}{2} \{\phi A_i^2 - B_i\} + \frac{\delta_3(\sigma^2)}{6} \{\phi^2 A_i^3 - 3\phi A_i B_i - C_i\} \right. \right. \\
&\quad \left. \left. + \frac{\delta_4(\sigma^2)}{24} \{\phi^3 A_i^4 - 6\phi^2 A_i^2 B_i - 4\phi A_i C_i + 3\phi B_i^2 - D_i\} \right] \right], \quad (2.4)
\end{aligned}$$

where $f_{ij}(y_{ij} | \theta_{ij}) = f_{ij}^*(y_{ij} | \theta_{ij}^*)|_{\theta_{ij}^* = \theta_{ij}}$ and

$$\begin{aligned} A_i &= \sum_{j=1}^{n_i} b_{ij}^{1/2} (y_{ij} - a'_{ij}), & B_i &= \sum_{j=1}^{n_i} b_{ij} a''_{ij}, \\ C_i &= \sum_{j=1}^{n_i} b_{ij}^{3/2} a'''_{ij}, & \text{and} & & D_i &= \sum_{j=1}^{n_i} b_{ij}^2 a^{\text{IV}}_{ij}, \end{aligned}$$

with a'_{ij} , a''_{ij} , a'''_{ij} and a^{IV}_{ij} the first, second, third, and the fourth order derivatives of $a(\theta_{ij}^*)$ in (2.1) with respect to θ_{ij}^* evaluated at $\theta_{ij}^* = \theta_{ij}$. In the GLMM setup, approximate likelihood functions similar to (2.4) have been used previously by other authors in different inference contexts. For example, we refer the reader to Sutradhar and Das (1995, p. 2690), Das and Sutradhar (1996, p. 475), and Lin (1997, pp. 311–312). The likelihood approximations used by these authors are, however, of order $o(\sigma^2)$, whereas the likelihood approximation in (2.4) is of order $o(\sigma^4)$.

In order to see how the unconditional means, variances, and covariances of n_i observations in the i th cluster under the GLMM setup are affected by the covariates and the random effects variation, we exploit the marginal likelihood $L_{i,j}$ and the bivariate marginal likelihood $L_{i,jk}$ in the following section to derive these moments. These first and second order moments are subsequently used to construct the MQL estimating equations for the regression effects, in the same section.

3. MQL ESTIMATING EQUATIONS FOR REGRESSION EFFECTS

The construction of the marginal estimating equations requires the mean vectors and covariance matrix of the responses in a cluster to be known. When these mean vectors and covariance matrices are too complicated to compute, some approximations to them are often used. In more complicated cases, sometimes a working covariance matrix is used for the true covariance matrix (cf. Zeger *et al.*, 1988). In the present approach, one can, however, compute these moments up to $o(\sigma^4)$ by using the joint density of the responses given in (2.4). The same joint density will be used in the next section to compute the third and the fourth order moments necessary to construct the MQL estimating equation for σ^2 . The computation of the moments up to order four will require the first eight moments of the exponential family distribution

$$f_{ij}(y_{ij} | \theta_{ij}) = \exp[\{y_{ij}\theta_{ij} - a(\theta_{ij})\} \phi + c(y_{ij}, \phi)]. \quad (3.1)$$

For convenience we provide these eight moments in the following lemma.

LEMMA 1. Let $m'_{ij,1} = E_{\exp}(Y_{ij})$ and $m_{ij,s} = E_{\exp}(Y_{ij} - m'_{ij,1})^s$ for $s=2, \dots, 8$, where, for example, $E_{\exp}(Y_{ij})$ denotes the expectation of Y_{ij} when y_{ij} has the exponential family pdf given by (3.1). Then these moments are given by

$$\begin{aligned} m'_{ij,1} &= a'_{ij}, & m_{ij,2} &= a''_{ij}/\phi, & m_{ij,3} &= a'''_{ij}/\phi^2, & m_{ij,4} &= a^{IV}_{ij}/\phi^3 + 3m_{ij,2}^2, \\ m_{ij,5} &= a^V_{ij}/\phi^4 + 10m_{ij,2}m_{ij,3}, \\ m_{ij,6} &= a^{VI}_{ij}/\phi^5 + 15m_{ij,2}m_{ij,4} + 10m_{ij,3}^2 - 30m_{ij,2}^3, \\ m_{ij,7} &= a^{VII}_{ij}/\phi^6 + 21m_{ij,2}m_{ij,5} + 35m_{ij,3}m_{ij,4} - 210m_{ij,2}^2m_{ij,3}, \end{aligned}$$

and

$$\begin{aligned} m_{ij,8} &= a^{VIII}_{ij}/\phi^7 + 28m_{ij,2}m_{ij,6} + 56m_{ij,3}m_{ij,5} - 630m_{ij,2}^2m_{ij,4} + 70m_{ij,2}^2m_{ij,4} \\ &\quad - 560m_{ij,2}m_{ij,3}^2 + 945m_{ij,2}^4, \end{aligned} \quad (3.2)$$

where $a'_{ij}, a''_{ij}, \dots, a^{VIII}_{ij}$, respectively, denote the first, second, ..., eighth order derivative of $a(\theta_{ij})$ in (3.1) with respect to θ_{ij} .

Note that the s th ($s=1, \dots, 8$) moment of y_{ij} in Lemma 1 is derived from the identity $(\int \partial^s f(y_{ij} | \theta_{ij}) / \partial \theta_{ij}^s dy_{ij} = 0$. We, however, do not show the computations for any of these eight moments. This is because the computations of these moments involve lengthy but straightforward algebras. Moreover, the first four of these moments are also available in any standard text book (cf. McCullagh and Nelder, 1989).

We now perform some basic integrations with respect to the exponential pdf (3.1) and provide the results in Lemmas 2 and 3. These integration results are expressed in terms of the moments given in Lemma 1, and they will be exploited to obtain the mean of the observed data in Theorem 1 and the variance and covariances in Theorem 2.

LEMMA 2. Let $h^{(r)}_{ij,(r+s)}$ denote the integral

$$h^{(r)}_{ij,(r+s)} = \int y_{ij}^r A_{i,j}^s f_{ij}(y_{ij} | \theta_{ij}) dy_{ij}, \quad (3.2)$$

where $A_{i,j} = b_{ij}^{1/2}(y_{ij} - a'_{ij})$, $f_{ij}(\cdot)$ is the exponential density as in (3.1), and r and s are nonnegative integers. Then for $r=1, 2$ and $s=0, 1, 2, 3, 4$, the h 's are given by

$$\begin{aligned} h^{(1)}_{ij,1} &= m'_{ij,1}, & h^{(1)}_{ij,2} &= b_{ij}^{1/2} m_{ij,2}, \\ h^{(1)}_{ij,(1+s)} &= b_{ij}^{s/2} \{m_{ij,(1+s)} + m_{ij,s} m'_{ij,1}\}, & \text{for } s &= 2, 3, 4, \end{aligned}$$

and

$$h_{ij,2}^{(2)} = m_{ij,2} + (m'_{ij,1})^2, \quad h_{ij,3}^{(2)} = b_{ij}^{1/2} \{m_{ij,3} + 2m_{ij,2}m'_{ij,1}\},$$

$$h_{ij,(2+s)}^{(2)} = b_{ij}^{s/2} \{m_{ij,(2+s)} + 2m_{ij,(1+s)}m'_{ij,1} + m_{ij,s}(m'_{ij,1})^2\},$$

for $s = 2, 3, 4$, where $m'_{ij,1}, m_{ij,2}, \dots, m_{ij,6}$ are given in Lemma 1.

The results in Lemma 2 are immediate from the direct integration (3.2).

Next, by integrating over $y_{ij'}$ for all $j' = 1, \dots, j-1, j+1, \dots, n_i$, the marginal density of the observed y_{ij} follows from (2.4) and is given by

$$L_{i,j}(\beta, \sigma^2) = f_{ij}(y_{ij} | \theta_{ij}) \left[1 + \phi \left[\frac{\sigma^2}{2} \{ \phi A_{i,j}^2 - B_{i,j} \} \right. \right. \\ \left. \left. + \frac{\delta_3(\sigma^2)}{6} \{ \phi^2 A_{i,j}^3 - 3\phi A_{i,j} B_{i,j} - C_{i,j} \} \right. \right. \\ \left. \left. + \frac{\delta_4(\sigma^2)}{24} \{ \phi^3 A_{i,j}^4 - 6\phi^2 A_{i,j}^2 B_{i,j} - 4\phi A_{i,j} C_{i,j} \right. \right. \\ \left. \left. + 3\phi B_{i,j}^2 - D_{i,j} \} \right] \right], \quad (3.3)$$

where $A_{i,j} = b_{ij}^{1/2}(y_{ij} - a'_{ij})$, $B_{i,j} = b_{ij} a''_{ij}$, $C_{i,j} = b_{ij}^{3/2} a'''_{ij}$, and $D_{i,j} = b_{ij}^2 a^{IV}_{ij}$. Note that further integration over (3.3) with respect to y_{ij} yields

$$\int L_{i,j}(\beta, \sigma^2) dy_{ij} = 1$$

which verifies that $L_i(\beta, \sigma^2)$ in (2.4) is a proper joint probability density function. The marginal density (3.3) is now exploited to derive the mean and the variance of Y_{ij} as in the following theorem.

THEOREM 1. For $j = 1, \dots, n_i$, let the observed response y_{ij} be generated following the marginal probability $L_{i,j}(\beta, \sigma^2)$ in (3.3). Then for h functions given as in Lemma 2,

$$E(Y_{ij}^r) = h_{ij,(r)}^{(r)} + \frac{\sigma^2}{2} d_{ij,(r+2)}^{(r)} + \frac{\delta_3(\sigma^2)}{2} d_{ij,(r+3)}^{(r)} + \frac{\delta_4(\sigma^2)}{24} d_{ij,(r+4)}^{(r)},$$

where

$$d_{ij,(r+2)}^{(r)} = \phi^2 h_{ij,(r+2)}^{(r)} - \phi h_{ij,(r)}^{(r)},$$

$$d_{ij,(r+3)}^{(r)} = \phi^3 h_{ij,(r+3)}^{(r)} - 3\phi^2 h_{ij,(r+1)}^{(r)} B_{i,j} - \phi h_{ij,(r)}^{(r)} C_{i,j},$$

and

$$d_{ij, (r+4)}^{(r)} = \phi^4 h_{ij, (r+4)}^{(r)} - 6\phi^3 h_{ij, (r+2)}^{(r)} B_{i \cdot j} - 4\phi^2 h_{ij, (r+1)}^{(r)} C_{i \cdot j} \\ + 3\phi^2 h_{ij, (r)}^{(r)} B_{i \cdot j}^2 - \phi h_{ij, (r)}^{(r)} D_{i \cdot j},$$

yielding the mean and the variance given by

$$M'_{ij, 1} = E(Y_{ij}) \quad (3.4)$$

and

$$M_{ij, 2} = E(Y_{ij}^2) - (M'_{ij, 1})^2, \quad (3.5)$$

respectively.

We now proceed to derive the covariance, $\text{cov}(Y_{ij}, Y_{ik})$, for $j \neq k$, $j, k = 1, \dots, n_i$. For this purpose, similar to Lemma 2, we perform some more basic integrations with respect to the exponential pdf (3.1) as in the following lemma.

LEMMA 3. For $j \neq k$, let $H_{ijk, (r+s+t)}^{(r,s)}$ denote the integral

$$H_{ijk, (r+s+t)}^{(r,s)} = \int y_{ij}^r y_{ik}^s A_{i, jk}^t f_{ij}(y_{ij} | \theta_{ij}) f_{ik}(y_{ik} | \theta_{ik}) dy_{ij} dy_{ik}, \quad (3.6)$$

where $f_{ij}(y_{ij} | \theta_{ij})$, for example, is the exponential pdf as in (3.1) and $A_{i, jk} = b_{ij}(y_{ij} - m'_{ij, 1}) + b_{ik}(y_{ik} - m'_{ik, 1})$ by (2.4), and r, s , and t are non-negative integers. Then, for $r = 1, s = 1$, and $t = 0, 1, \dots, 4$, the H functions are given by

$$H_{ijk, (r+s+t)}^{(r,s)} = \sum_{u=1}^{t+1} {}^t C_{u-1} h_{ij, (t+2-u)}^{(r)} h_{ik, u}^{(s)}, \quad (3.7)$$

where h functions are as in Lemma 2 and ${}^t C_{u-1}$ denotes the number of ways that $u-1$ functions can be chosen from t functions.

The results of Lemma 3 are next exploited to derive the covariance between Y_{ij} and Y_{ik} as in the following theorem.

THEOREM 2. For $j \neq k, j, k = 1, \dots, n_i$, let the observed response y_{ij} and y_{ik} be generated following the bivariate pdf $L_{i, jk}$ obtained from (2.4). Then, for $r = s = 1$,

$$E(Y_{ij}^r Y_{ij}^s) = H_{ijk, (r+s)}^{(r,s)} + \frac{\sigma^2}{2} d_{ijk, (r+s+2)} + \frac{\delta_3(\sigma^2)}{6} d_{ijk, (r+s+3)}^{(r,s)} + \frac{\delta_4(\sigma^2)}{24} d_{ijk, (r+s+4)}^{(r,s)}, \quad (3.8)$$

where for $u = 2, 3, 4$, $d_{ijk, (r+s+u)}^{(r,s)}$ are obtained from $d_{ij, (r+s+u)}^{(r)}$ in Theorem 1, by replacing $h_{ij, (r+s+u)}^{(r)}$, $B_{i,j}$, and $C_{i,j}$ with $H_{ijk, (r+s+u)}^{(r,s)}$, $B_{i,jk}$, and $C_{i,jk}$, respectively, yielding the covariance given by

$$M_{ijk, 2} = E(Y_{ij} Y_{ik}) - M'_{ij, 1} M'_{ik, 1} \quad (3.9)$$

where $M'_{ij, 1}$ and $M'_{ik, 1}$ are given in Theorem 1. In (3.8), $B_{i,jk} = b_{ij} a_{ij}'' + b_{ik} a_{ik}''$ and $C_{i,jk} = b_{ij}^{3/2} a_{ij}''' + b_{ik}^{3/2} a_{ik}'''$ by (2.4).

Note that after some more lengthy algebras, one may show that the mean and the variance of Y_{ij} given in Theorem 1 reduces to

$$M'_{ij, 1} = a'_{ij} + \frac{\sigma^2}{2} b_{ij} a_{ij}''' + \frac{\delta_3(\sigma^2)}{6} b_{ij}^{3/2} a_{ij}^{\text{IV}} + \frac{\delta_4(\sigma^2)}{24} b_{ij}^2 a_{ij}^{\text{V}} \quad (3.10)$$

and

$$\begin{aligned} M_{ijj, 2} = & \left[(a'_{ij})^2 + \sigma^2 b_{ij} \{a'_{ij} a_{ij}''' + (a_{ij}'')^2\} + \frac{\delta_3(\sigma^2)}{3} b_{ij}^{3/2} \{a'_{ij} a_{ij}^{\text{IV}} \right. \\ & \left. + 3a_{ij}'' a_{ij}'''\} + \frac{\delta_4(\sigma^2)}{12} b_{ij}^2 \{a'_{ij} a_{ij}^{\text{V}} + 4a_{ij}'' a_{ij}^{\text{IV}} + 3(a_{ij}''')^2\} \right] \\ & + \frac{1}{\phi} \left[a_{ij}'' + \frac{\sigma^2}{2} b_{ij} a_{ij}^{\text{IV}} + \frac{\delta_3(\sigma^2)}{6} b_{ij}^{3/2} a_{ij}^{\text{V}} + \frac{\delta_4(\sigma^2)}{24} b_{ij}^2 a_{ij}^{\text{VI}} \right] - (M'_{ij, 1})^2, \end{aligned} \quad (3.11)$$

respectively. Similarly, the covariance given in Theorem 2 may be simplified as

$$\begin{aligned} M_{ijk, 2} = & a'_{ij} a'_{ik} + \frac{\sigma^2}{2} \left[b_{ij} a_{ij}''' a'_{ik} + 2b_{ij}^{1/2} b_{ik}^{1/2} a_{ij}'' a_{ik}'' + b_{ik} a'_{ij} a_{ik}''' \right] \\ & + \frac{\delta_3(\sigma^2)}{6} [b_{ij}^{3/2} a_{ij}^{\text{IV}} a'_{ik} + 3b_{ij} b_{ik}^{1/2} a_{ij}''' a_{ik}'' + 3b_{ij}^{1/2} b_{ik} a_{ij}'' a_{ik}''' + b_{ik}^{3/2} a'_{ij} a_{ik}^{\text{IV}}] \\ & + \frac{\delta_4(\sigma^2)}{24} [b_{ij}^2 a_{ij}^{\text{V}} a'_{ik} + 4b_{ij}^{3/2} b_{ik}^{1/2} a_{ij}^{\text{IV}} a_{ik}'' + 6b_{ij} b_{ik} a_{ij}''' a_{ik}''' \\ & + 4b_{ij}^{1/2} b_{ik}^{3/2} a_{ij}'' a_{ik}^{\text{IV}} + b_{ik}^2 a'_{ij} a_{ik}^{\text{V}}] - M'_{ij, 1} M'_{ik, 1}. \end{aligned} \quad (3.12)$$

The mean, variance, and covariance given in (3.10)–(3.12) are correct up to order σ^5 for asymmetric γ 's and they are correct up to order σ^6 for symmetric γ 's. Further, note that the formulas for the mean (3.10), variance (3.11), and covariance (3.12) are expressed directly in terms of the notations or functions used in the original likelihood function in (2.4). But similar simplification may not be worth doing for the third and the fourth order moments of the responses. For these higher order moments we will follow the unified notations used in Theorems 1 and 2. The formulas for the third and the fourth order moments and the construction of the MQL estimating equation for σ^2 are given in the next section.

Turning back to the MQL estimation for the regression parameter β , let

$$M'_{i,1} = [M'_{i1,1}, \dots, M'_{ij,1}, \dots, M'_{in_i,1}]^T \quad (3.13)$$

be the $n_i \times 1$ mean vector for n_i observations in the i th ($i = 1, \dots, k$) cluster, where $M'_{ij,1}$ for all $j = 1, \dots, n_i$ is given by (3.4) or (3.10). Also, let

$$M_{i,2} = (M_{ijk,2}) \quad (3.14)$$

be the $n_i \times n_i$ covariance matrix of $Y_i = [Y_{i1}, \dots, Y_{ij}, \dots, Y_{in_i}]^T$ under the GLMM, where for $j = k = 1, \dots, n_i$ the diagonal elements $M_{ijj,2}$ are computed from (3.5) or (3.11) and for $j \neq k$, the off-diagonal elements $M_{ijk,2}$ are computed from (3.9) or equivalently (3.12). Then for known σ^2 , the MQL estimator $\hat{\beta}_{\text{MQL}}$, as discussed by McCullagh (1983) (see also Wedderburn, 1974) is the root of the equation

$$\sum_{i=1}^I D_i^T M_{i,2}^{-1} (y_i - M'_{i,1}) = 0, \quad (3.15)$$

where $D_i = \partial\{M'_{i,1}\}/\partial\beta$ is the $n_i \times p$ first derivative matrix of $M'_{i,1}$ with respect to β , which is given in Appendix A. The solution of (3.15), that is, $\hat{\beta}_{\text{MQL}}$, may be obtained by the customary Newton–Raphson method. Given the value $\hat{\beta}_{\text{MQL}}(t)$ at the t th iteration, $\hat{\beta}_{\text{MQL}}(t+1)$ is obtained as

$$\hat{\beta}_{\text{MQL}}(t+1) = \hat{\beta}_{\text{MQL}}(t) + \left[\sum_{i=1}^I D_i^T M_{i,2}^{-1} D_i \right]_t^{-1} \left[\sum_{i=1}^I D_i^T M_{i,2}^{-1} (y_i - M'_{i,1}) \right]_t, \quad (3.16)$$

where $[\cdot]_t$ denotes that the expression within the brackets is evaluated at $\hat{\beta}_{\text{MQL}}(t)$. Note that for known σ^2 and suitable initial values for β (small positive or negative), the iterative algorithm (3.16) would converge rapidly as the $M'_{i,1}$ in (3.16) is a smooth (cf. Seber and Wild, 1989, p. 599–600) mean function of β under the exponential family distribution.

Next, it follows under some mild regularity conditions that $I^{12}(\hat{\beta}_{\text{MQL}} - \beta)$ is asymptotically multivariate normal with zero mean vector and covariance matrix V_{β} , which may be consistently estimated by

$$\hat{V}_{\beta} = \lim_{I \rightarrow \infty} I \left[\sum_{i=1}^I D_i^T M_{i,2}^{-1} D_i \right]_{\hat{\beta}_{\text{MQL}}}^{-1}. \quad (3.17)$$

Note that $\hat{\beta}_{\text{MQL}}$ is a consistent estimator for β . Further, as we have used the almost correct covariance matrix in the estimating equation (3.15), this MQL estimate will be naturally more efficient than any other traditional MQL estimate computed by using the so-called working covariance matrix for the true covariance.

The computations of the MQL estimate $\hat{\beta}_{\text{MQL}}$ by (3.16) and of its covariance estimate by (3.17) are relatively much simpler as compared to the estimation of the likelihood estimate of β , say, $\hat{\beta}_{\text{MLE}}$, and its covariance estimate. This is because the likelihood estimate $\hat{\beta}_{\text{MLE}}$ requires the computation of the second derivative of the likelihood function $\prod_{i=1}^I L_i(\beta, \sigma^2)$ (2.4), with respect to β , and its covariance estimate requires the computation of the Fisher information matrix, which is quite involved under the GLMM setup.

3.1. Effect of the Proposed MQL Estimation for Regression Coefficient: A Numerical Illustration

In order to examine the effect of using the proposed MQL estimation for the regression parameters discussed in the paper, we make a numerical comparison of the variances of the regression estimates computed by using (3.17) with those of the existing methods, such as the MQL approach discussed in Breslow and Clayton (1993, Section 3). Without any loss of generality, we consider a binary mixed model with normally distributed random effects. More specifically, in our notation, $a'(\theta_{ij}) = 1/[1 + \exp(-\theta_{ij})]$, with $\theta_{ij} = x_{ij}^T \beta$. Following Zeger *et al.* (1988), Breslow and Clayton (1993) approximated the mean vector $M'_{i,1}$ by $p_i^* = (p_{i1}^*, \dots, p_{ij}^*, \dots, p_{in_i}^*)^T$, where

$$p_{ij}^* = 1/[1 + \exp\{-c_{ij}^T x_{ij}^T \beta\}], \quad (3.18)$$

with $c_{ij} = (1 + c^2 \sigma^2)^{-1/2}$ and $c = 16(3^{1/2})/15\pi$. Similarly, the covariance matrix of y_i was approximated by

$$\Sigma_i^* = V_{i0} + \sigma^2 V_{i0} U_{n_i} V_{i0}, \quad (3.19)$$

where U_{n_i} is the $n_i \times n_i$ unit matrix and $V_{i0} = \text{diag}[p_{i1}(\gamma_i = 0) q_{i1}(\gamma_i = 0), \dots, p_{in_i}(\gamma_i = 0) q_{in_i}(\gamma_i = 0)]$, with $p_{ij}(\gamma_i) = 1/[1 + \exp\{-x_{ij}^T \beta - \gamma_i\}]$. Let $\hat{\beta}_{\text{MQL1}}$ denote the estimate of β in this approach. Similar to (3.17), it then follows

that $I^{1/2}(\hat{\beta}_{\text{MQL1}} - \beta)$ is asymptotically multivariate normal with zero mean vector and covariance matrix V_{β}^* , which may be consistently estimated by

$$\hat{V}_{\beta}^* = \lim_{I \rightarrow \infty} I \left[\sum_{i=1}^I P_i^* \Sigma_i^{*-1} P_i^{*T} \right]_{\hat{\beta}_{\text{MQL1}}}^{-1}, \quad (3.20)$$

where $P_i^* = X_i^T M_i^* C_i^*$, with $M_i^* = \text{diag}[p_{i1}^* q_{i1}^*, \dots, p_{ij}^* q_{ij}^*, \dots, p_{in_i}^* q_{in_i}^*]$ and $C_i^* = \text{diag}[c_{i1}, \dots, c_{in_i}]$.

Now to examine the efficiency loss due to using the existing MQL approach, we compute the relative efficiency (ratio of the variances) of the u th ($u = 1, \dots, p$) regression component as

$$\text{reff}(\hat{\beta}_{u(\text{MQL1})}) = v_{\text{MQL2}}(u, u) / v_{\text{MQL1}}(u, u), \quad (3.21)$$

where $v_{\text{MQL2}}(u, u)$ and $v_{\text{MQL1}}(u, u)$ are the u th diagonal elements of the covariance matrices, $\text{cov}(\hat{\beta}_{\text{MQL2}})$ (3.17) and $\text{cov}(\hat{\beta}_{\text{MQL1}})$ (3.20), respectively. Note that here we have used $\hat{\beta}_{\text{MQL2}}$ for the new MQL regression estimator computed by (3.16). In order to see how relative efficiency can vary with regard to the change in σ^2 values, we have computed the relative efficiency of $\hat{\beta}_{\text{MQL1}}$ by (3.21) for two design matrices with $n_i = 6$ and $p = 2$ for $i = 1, \dots, 100$. The two covariates under the first design (D_1) were chosen (cf. Liang and Zeger, 1986) as

$$\begin{aligned} x_{ij1} &= 1 & \text{for } j = 1, \dots, 6; \quad i = 1, \dots, 100; \\ x_{ij2} &= 1/j & \text{for } j = 1, \dots, 6; \quad i = 1, \dots, 100; \end{aligned}$$

and under the second design (D_2) they were

$$\begin{aligned} x_{ij1} &= 1 & \text{for } j = 1, \dots, 6; \quad i = 1, \dots, 100; \\ x_{ij2} &= \begin{cases} -1 & \text{for } j = 1, \dots, 3; \quad i = 1, \dots, 50 \\ 0 & \text{for } j = 4, \dots, 6; \quad i = 1, \dots, 50 \\ -1 & \text{for } j = 1, 2; \quad i = 51, \dots, 100 \\ 0 & \text{for } j = 3, 4; \quad i = 51, \dots, 100 \\ 1 & \text{for } j = 5, 6; \quad i = 51, \dots, 100. \end{cases} \end{aligned}$$

The relative efficiencies are reported in Table I. It is clear from the table that although the efficiency loss is negligible for small values of $\sigma^2 \leq 0.3$, the relative efficiency may, however, be quite low such as 72% for the intercept parameter and 87% for the slope parameter, for $\sigma^2 = 0.9$ under D_2 . Under both designs, the relative efficiencies of the regression estimators appear to get smaller as σ^2 gets larger, the situation being worse under D_2 as

TABLE I

Percentage Relative Efficiency of $\hat{\beta}_{\text{MQL1}} = (\hat{\beta}_{1(\text{MQL1})}, \hat{\beta}_{2(\text{MQL1})})^T$ to the Proposed MQL Estimator $\hat{\beta}_{\text{MQL2}} = (\hat{\beta}_{1(\text{MQL2})}, \hat{\beta}_{2(\text{MQL2})})^T$ for Selected Values of σ^2 and β_1, β_2

Design	Regression Coefficient	Relative Efficiency of	Values of σ^2					
			0.10	0.20	0.30	0.50	0.70	0.90
D_1	$\beta_1 = 1, \beta_2 = -1$	$\hat{\beta}_1(\text{MQL1})$	99	98	97	95	94	93
		$\hat{\beta}_2(\text{MQL1})$	99	98	97	95	93	91
	$\beta_1 = 0.25, \beta_2 = 0.25$	$\hat{\beta}_1(\text{MQL1})$	99	98	96	93	88	82
		$\hat{\beta}_2(\text{MQL1})$	99	98	97	94	92	89
	$\beta_1 = 0.25, \beta_2 = -0.25$	$\hat{\beta}_1(\text{MQL1})$	99	98	97	92	87	81
		$\hat{\beta}_2(\text{MQL1})$	99	98	97	94	91	87
D_2	$\beta_1 = 1, \beta_2 = -1$	$\hat{\beta}_1(\text{MQL1})$	99	99	98	97	96	95
		$\hat{\beta}_2(\text{MQL1})$	99	99	98	97	95	92
	$\beta_1 = 0.25, \beta_2 = 0.25$	$\hat{\beta}_1(\text{MQL1})$	99	97	95	89	81	73
		$\hat{\beta}_2(\text{MQL1})$	99	98	97	94	91	88
	$\beta_1 = 0.25, \beta_2 = -0.25$	$\hat{\beta}_1(\text{MQL1})$	99	97	94	88	81	72
		$\hat{\beta}_2(\text{MQL1})$	99	98	97	94	91	87

compared to D_1 for the intercept parameter. Under both designs, the efficiency loss appears to be significant even for moderate values of σ^2 such as $\sigma^2 = 0.5$ and 0.7 . These relative efficiency results, therefore, indicate that the proposed MQL approach leads to better regression estimates as compared to the existing MQL approaches, such as the MQL approach discussed in Breslow and Clayton (1993).

4. MQL ESTIMATING EQUATION FOR VARIANCE COMPONENT

In the GLMM setup, the response variable is subject to overdispersion which is caused by the variance component of the random effects. This overdispersion parameter does not only influence the marginal variance, it may also influence the mean of the response variable. For example, consider the Poisson case with

$$\begin{aligned} a(\theta_{ij}^*) &= \exp(\theta_{ij}^*) \\ &= \exp \{ \theta_{ij} + b_{ij}^{1/2} \gamma_i \}. \end{aligned}$$

For this case, the unconditional mean of Y_{ij} given in (3.10) reduces to

$$M'_{ij,1} = e^{\theta_{ij}} \left[1 + \frac{\sigma^2}{2} b_{ij} + \frac{\delta_3(\sigma^2)}{6} b_{ij}^{3/2} + \frac{\delta_4(\sigma^2)}{24} b_{ij}^2 \right],$$

which, for $b_{ij} = 1$ and normal $\delta_3(\sigma^2) = 0$, $\delta_4(\sigma^2) = 3\sigma^4$, further reduces to

$$\begin{aligned} M'_{ij,1} &= e^{\theta_{ij}} \left[1 + \frac{\sigma^2}{2} + \frac{1}{2!} \left(\frac{\sigma^2}{2} \right)^2 \right] \\ &\simeq e^{\theta_{ij} + \sigma^2}, \end{aligned} \quad (4.1)$$

provided that σ^6 is negligible. The mean function in (4.1) clearly depends on the overdispersion parameter σ^2 . Similarly, the mean function of the binary mixed variable will also depend on the σ^2 parameter. This observation motivates one to use both the first and the second order moments while estimating the variance component of the random effects by using the MQL approach.

Let u_{i1}^* and u_{i2}^* denote the vectors of squared and pairwise product responses of the i th cluster. That is, $u_{i1}^* = [y_{i1}^2, \dots, y_{ij}^2, \dots, y_{in_i}^2]^T$ and $u_{i2}^* = [y_{i1}y_{i2}, \dots, y_{i,n_i-1}y_{in_i}]^T$. These vectors are of dimension $n_i \times 1$ and $\{n_i(n_i-1)/2\} \times 1$ respectively. Further, let

$$m_{i1,2}^* = [M'_{i1,2}, \dots, M'_{ij,2}, \dots, M'_{in_i,2}]^T,$$

and

$$m_{i2,2}^* = [M'_{i12,2}, \dots, M'_{ijk,2}, \dots, M'_{i,n_i-1,n_i,2}]^T,$$

where for $j = 1, \dots, n_i$, $M'_{ij,2} = E(Y_{ij}^2)$ are given as in Theorem 1, and for $j \neq k$, $j, k = 1, \dots, n_i$, $M'_{ijk,2} = E(Y_{ij}Y_{ik})$ are given as in Theorem 2. As, all three vectors of first and second order moments, namely, $M'_{i,1}$ in (3.13), and $m_{i1,2}^*$, and $m_{i2,2}^*$, contain the variance component σ^2 , we now write the MQL estimating equation for σ^2 as

$$\sum_{i=1}^I D_i^{*T} M_i^{*-1} f_i = 0, \quad (4.2)$$

where

$$\begin{aligned} f_i &= [(y_i - M'_{i,1})^T, (u_{i1}^* - m_{i1,2}^*)^T, (u_{i2}^* - m_{i2,2}^*)^T]^T, \\ D_i^* &= \partial [M_{(4.7)i,1}^{*T}, m_{i1,2}^{*T}, m_{i2,2}^{*T}]^T / \partial \sigma^2, \end{aligned}$$

and

$$M_i^* = \begin{bmatrix} \text{var}(y_i) & \text{cov}(y_i, u_{i1}^*) & \text{cov}(y_i, u_{i2}^*) \\ & \text{var}(u_{i1}^*) & \text{cov}(u_{i1}^*, u_{i2}^*) \\ & & \text{var}(u_{i2}^*) \end{bmatrix},$$

with $\text{var}(y_i) = M_{i,2}$ as in (3.14). The computations of the other submatrices in M_i^* are discussed below, and the computation for the D_i^* matrix is shown in Appendix B.

Let $\hat{\sigma}_{\text{MQL}}^2$ be the solution of (4.2). This solution may be obtained by the customary Newton–Raphson method. Given the value of $\hat{\sigma}_{\text{MQL}}^2(t)$ at the t th iteration, similar to (3.16), $\hat{\sigma}_{\text{MQL}}^2(t+1)$ is obtained as

$$\hat{\sigma}_{\text{MQL}}^2(t+1) = \hat{\sigma}_{\text{MQL}}^2(t) + \left[\sum_{i=1}^I D_i^{*\top} M_i^{*-1} D_i^* \right]_t^{-1} \left[\sum_{i=1}^I D_i^{*\top} M_i^{*-1} f_i \right]_t, \quad (4.3)$$

where $[\cdot]_t$ denotes that the expression within the brackets is evaluated at $\hat{\sigma}_{\text{MQL}}^2(t)$. Further, it follows that $I^{1/2}(\hat{\sigma}_{\text{MQL}}^2 - \sigma^2)$ is asymptotically (as $I \rightarrow \infty$) univariate normal with zero mean and variance V_{σ^2} , which may be consistently estimated by

$$\hat{V}_{\sigma^2} = \lim_{I \rightarrow \infty} I \left[\sum_{i=1}^I D_i^{*\top} M_i^{*-1} D_i^* \right]_{\hat{\sigma}_{\text{MQL}}^2}^{-1}, \quad (4.4)$$

where $[\cdot]_{\hat{\sigma}_{\text{MQL}}^2}$ denotes that the expression within the bracket is evaluated at $\hat{\sigma}_{\text{MQL}}^2$. Note that for given β , the iterative algorithm (4.3) for σ^2 converges at a slower rate as compared to that for β in (3.16). This is because the variance function in f_i may not be smooth as compared to the mean function in (3.16). The convergence rate, however, may be improved by choosing an appropriate initial value for the variance component. For example, Breslow and Clayton (1993, Section 4.1) set a small positive initial value for σ^2 . Alternatively, as the function involved in (4.2) is a scalar function in σ^2 , one may search for an initial value of σ^2 by computing the function for many possible values of σ^2 and choosing the initial value for which the value of the function is close to zero.

We now turn back to the construction of the covariance matrix M_i^* for which only $\text{var}(y_i) = M_{i,2}$ is known so far by (3.14). We construct the submatrices in the sequence of $\text{cov}(y_i, u_{i1}^*)$, $\text{cov}(y_i, u_{i2}^*)$, $\text{var}(u_{i1}^*)$, $\text{cov}(u_{i1}^*, u_{i2}^*)$, and $\text{var}(u_{i2}^*)$.

The computation of $\text{cov}(y_i, u_{i1}^*)$ requires $E(Y_{ij}^3)$ for all $j = 1, \dots, n_i$ and $E(Y_{ij} Y_{ik}^2)$ for all $j \neq k$, $j, k = 1, \dots, n_i$. Following Lemma 2, for $r = 3$, and $s = 0, 1, 2, 3, 4$, we obtain the integration results $h_{ij, (r+s)}^{(r)}$ as in the following lemma. For convenience, we also provide the integration results for $r = 4$ in the same lemma. The latter results will be used to compute $E(Y_{ij}^4)$.

LEMMA 4. *Let $h_{ij, (r+s)}^{(r)}$ be the integral (3.2) as in Lemma 2. Then for $r = 3, 4$ and $s = 0, 1, 2, 3, 4$, the h 's are given by*

$$h_{ij,3}^{(3)} = m_{ij,3} + 3m_{ij,2}m'_{ij,1} + (m'_{ij,1})^3,$$

$$h_{ij,4}^{(3)} = b_{ij}^{1/2} [m_{ij,4} + 3m_{ij,3}m'_{ij,1} + 3m_{ij,2}(m'_{ij,1})^2],$$

$$h_{ij,(3+s)}^{(3)} = b_{ij}^{s/2} [m_{ij,(3+s)} + 3m_{ij(2+s)}m'_{ij,1} + m_{ij,(1+s)}(m'_{ij,1})^2 + m_{ij,s}(m'_{ij,1})^3],$$

for $s = 2, 3, 4$; and

$$h_{ij,4}^{(4)} = m_{ij,4} + 4m_{ij,3}m'_{ij,1} + 6m_{ij,2}(m'_{ij,1})^2 + (m'_{ij,1})^4,$$

$$h_{ij,5}^{(4)} = m_{ij,5} + 4m_{ij,4}m'_{ij,1} + 6m_{ij,2}(m'_{ij,1})^3,$$

$$h_{ij,(4+s)}^{(4)} = b_{ij}^{s/2} [m_{ij,(4+s)} + 4m_{ij,(3+s)}m'_{ij,1} + 6m_{ij,(2+s)}(m'_{ij,1})^2 + 4m_{ij,(1+s)}(m'_{ij,1})^3 + m_{ij,s}(m'_{ij,1})^4],$$

for $s = 2, 3, 4$, where $m'_{ij,1}, m_{ij,2}, \dots, m_{ij,s}$ are given in Lemma 1.

By using Lemma 4, we now obtain the formulas for $E(Y_{ij}^3)$ and $E(Y_{ij}Y_{ik}^2)$ as in Theorem 3.

THEOREM 3. For $j = 1, \dots, n_i$, the third order moments $E(Y_{ij}^3)$ may be obtained from Theorem 1 by putting $r=3$ in the formula for $E(Y_{ij}^r)$, and similarly for $j \neq k$, $j, k = 1, \dots, n_i$, $E(Y_{ij}Y_{ik}^2)$ may be computed from Theorem 2 by using $r=1$ and $s=2$ in the formula for $E(Y_{ij}^rY_{ik}^s)$, where for $u = 0, 1, \dots, 4$, $h_{ij,(3+u)}^{(3)}$ and $h_{ij,(4+u)}^{(4)}$ are now given by Lemma 4.

Let $M_{i,3}^{(1)} = (M_{ijk,3}^{(1)})$ denote the $n_i \times n_i$ covariance matrix $\text{cov}(y_i, u_{i1}^*)$. Then the diagonal and off-diagonal elements of this matrix may be computed as

$$M_{ij,3}^{(1)} = E(Y_{ij}^3) - E(Y_{ij})E(Y_{ij}^2), \quad \text{for } j = 1, \dots, n_i, \quad (4.5)$$

and

$$M_{ijk,3}^{(1)} = E(Y_{ij}Y_{ik}^2) - E(Y_{ij})E(Y_{ik}^2), \quad \text{for } j \neq k, j, k = 1, \dots, n_i,$$

respectively, where $E(Y_{ij})$ and $E(Y_{ik}^2)$ are given by (3.4) and (3.5) in Theorem 1.

Next to compute the elements of the covariance matrix $\text{cov}(y_i, u_{i2}^*) = M_{i,3}^{(2)} = (M_{ijk,3}^{(2)})$, say, we require to compute

$$E(Y_{ij}^2Y_{ik}), E(Y_{ij}Y_{ik}Y_{il}), \quad \text{and} \quad E(Y_{ij}Y_{ik}^2),$$

for $j \neq k \neq l$. The formulas for $E(Y_{ij}Y_{ik}^2)$ are shown in Theorem 3. The computation for $E(Y_{ij}^2Y_{ik})$ may be carried out in the manner similar to that of $E(Y_{ij}Y_{ik}^2)$ by using $r=2$ and $s=1$ in the formula for $E(Y_{ij}^rY_{ik}^s)$ in Theorem 2. Now, the formula for $E(Y_{ij}Y_{ik}Y_{il})$ is given in Theorem 4.

THEOREM 4. For $j \neq k \neq l$, $j = k = l = 1, \dots, n_i$, let the observed responses y_{ij} , y_{ik} and y_{il} be generated following the trivariate joint pdf $L_{i,jkl}$ obtained from (2.4). Then, for positive integers r , s , and t ,

$$\begin{aligned} E(Y_{ij}^r Y_{ik}^s Y_{il}^t) &= H_{ijkl, (r+s+t)}^{(r,s,t)} + \frac{\sigma^2}{2} D_{ijkl, (r+s+t+2)}^{(r,s,t)} \\ &+ \frac{\delta_3(\sigma^2)}{6} D_{ijkl, (r+s+t+3)}^{(r,s,t)} + \frac{\delta_4(\sigma^2)}{24} D_{ijkl, (r+s+t+4)}^{(r,s,t)}, \end{aligned} \quad (4.6)$$

where

$$\begin{aligned} D_{ijkl, (r+s+t+2)}^{(r,s,t)} &= \phi^2 H_{ijkl, (r+s+t+2)}^{(r,s,t)} - \phi H_{ijkl, (r+s+t)}^{(r,s,t)} B_{i,jkl}, \\ D_{ijkl, (r+s+t+3)}^{(r,s,t)} &= \phi^3 H_{ijkl, (r+s+t+3)}^{(r,s,t)} - 3\phi^2 H_{ijkl, (r+s+t+1)}^{(r,s,t)} B_{i,jkl} \\ &\quad - \phi H_{ijkl, (r+s+t)}^{(r,s,t)} C_{i,jkl}, \end{aligned}$$

and

$$\begin{aligned} D_{ijkl, (r+s+t+4)}^{(r,s,t)} &= \phi^4 H_{ijkl, (r+s+t+4)}^{(r,s,t)} - 6\phi^3 H_{ijkl, (r+s+t+2)}^{(r,s,t)} B_{i,jkl} \\ &\quad - 4\phi^2 H_{ijkl, (r+s+t+1)}^{(r,s,t)} C_{i,jkl} + 3\phi^2 H_{ijkl, (r+s+t)}^{(r,s,t)} B_{i,jkl}^2 \\ &\quad - \phi H_{ijkl, (r+s+t)}^{(r,s,t)} D_{i,jkl}, \end{aligned}$$

with

$$\begin{aligned} &H_{ijkl, (r+s+t+u)}^{(r,s,t)} \\ &= \int y_{ij}^r y_{ik}^s y_{il}^t A_{i,jkl}^u f_{ij}(y_{ij} | \theta_{ij}) f_{ik}(y_{ik} | \theta_{ik}) f_{il}(y_{il} | \theta_{il}) dy_{ij} dy_{ik} dy_{il}, \end{aligned} \quad (4.7)$$

where for $u = 0, 1, 2, 3$, and 4 , the formulas for H functions are given in the Appendix C.

Now, based on Theorems 1, 2 and 3, we can compute all the elements of the covariance matrix $M_{i,3}^{(2)}$ by using the formulas

$$\begin{aligned} \text{cov}(Y_{ij}, Y_{ik}^2) &= E(Y_{ij} Y_{ik}^2) - E(Y_{ij}) E(Y_{ik}^2), \\ \text{cov}(Y_{ij}, Y_{ik} Y_{il}) &= E(Y_{ij} Y_{ik} Y_{il}) - E(Y_{ij}) E(Y_{ik} Y_{il}), \end{aligned} \quad (4.8)$$

and

$$\text{cov}(Y_{ij}^2, Y_{ik}) = E(Y_{ij}^2 Y_{ik}) - E(Y_{ij}^2) E(Y_{ik}).$$

Next, we turn to the computation of the variance matrix $M_{i,4}^{(1)} = \text{var}(u_{i1}^*)$. For this, one requires to compute $E(Y_{ij}^4)$ and $E(Y_{ij}^2 Y_{ik}^2)$ which are of fourth order. By using $r=4$, it follows from Theorem 1 that

$$E(Y_{ij}^4) = h_{ij,(4)}^{(4)} + \frac{\sigma^2}{2} d_{ij,(6)}^{(4)} + \frac{\delta_3(\sigma^2)}{2} d_{ij,(7)}^{(4)} + \frac{\delta_4(\sigma^2)}{24} d_{ij,(8)}^{(4)}, \quad (4.9)$$

which requires the values of the h functions ranging from $h_{ij,(4)}^{(4)}$ to $h_{ij,(8)}^{(4)}$. These h -functions are available in Lemma 4. Similarly, by putting $r=2$ and $s=2$ in (3.8), it follows from Theorem 2 that

$$E(Y_{ij}^2 Y_{ik}^2) = H_{ijk,(4)}^{(2,2)} + \frac{\sigma^2}{2} d_{ijk,(6)}^{(2,2)} + \frac{\delta_3}{6} d_{ijk,(7)}^{(2,2)} + \frac{\delta_4(\sigma^2)}{24} d_{ijk,(8)}^{(2,2)}, \quad (4.10)$$

where $d_{ijk,(2+2+t)}^{(2,2)}$ for $t=0, \dots, 4$ are computed from the expression given in (3.8). Consequently, we obtain the diagonal and off-diagonal elements of the matrix $M_{i,4}^{(1)}$ as

$$M_{ij,4}^{(1)} = E(Y_{ij}^4) - \{E(Y_{ij}^2)\}^2$$

and

$$M_{ijk,4}^{(1)} = E(Y_{ij}^2 Y_{ik}^2) - E(Y_{ij}^2) E(Y_{ik}^2),$$

for $j=1, \dots, n_i$ and $j \neq k, j, k=1, \dots, n_i$, respectively. In (4.11), $E(Y_{ij}^2)$ is computed from Theorem 1 for all $j=1, \dots, n_i$.

In order to compute the covariance matrix $M_{i,4}^{(2)} = \text{cov}(u_{i1}^*, u_{i2}^*)$, we require $E(Y_{ij}^3 Y_{ik})$, $E(Y_{ij}^2 Y_{ik} Y_{il})$ for $j \neq k \neq l$. Here $E(Y_{ij}^3 Y_{ik})$ is computed from Theorem 2 by using $r=3$ and $s=1$. The latter fourth order moments are computed from Theorem 4 by using $r=2$, $s=1$, and $l=1$. Hence, the elements of the covariance matrix $\text{cov}(u_{i1}^*, u_{i2}^*)$ are computed as

$$\begin{aligned} \text{cov}(Y_{ij}^2, Y_{ij} Y_{ik}) &= E(Y_{ij}^3 Y_{ik}) - E(Y_{ij}^2) E(Y_{ij} Y_{ik}), \\ \text{cov}(Y_{ij}^2, Y_{ik} Y_{il}) &= E(Y_{ij}^2 Y_{ik} Y_{il}) - E(Y_{ij}^2) E(Y_{ik} Y_{il}), \end{aligned} \quad (4.12)$$

and consequently the $M_{i,4}^{(2)}$ covariance matrix is computed.

Next, we compute the last covariance matrix $\text{var}(u_{i2}^*) = M_{i,4}^{(3)}$, say. The diagonal elements of this matrix are computed as

$$\text{cov}(Y_{ij} Y_{ik}, Y_{ij} Y_{ik}) = E(Y_{ij}^2 Y_{ik}^2) - \{E(Y_{ij} Y_{ik})\}^2, \quad (4.13)$$

where $E(Y_{ij}^2 Y_{ik}^2)$ is computed by (4.10) and $E(Y_{ij} Y_{ik})$ is computed by (3.9) in Theorem 2. The construction of the off-diagonal elements requires us to compute

$$E(Y_{ij} Y_{ik} Y_{il} Y_{im}), E(Y_{ij}^2 Y_{ik} Y_{il}), E(Y_{ij} Y_{ik}^2 Y_{il}), \quad \text{and} \quad E(Y_{ij} Y_{ik} Y_{il}^2).$$

The last three expectations may be derived from Theorem 4, by exploiting $r = 2, s = 1, t = 1$; $r = 1, s = 2, t = 1$; and $r = 1, s = 1, t = 2$, respectively. The first expectation $E(Y_{ij}Y_{ik}Y_{il}Y_{im})$ may be computed from the general results given in Theorem 5 below.

THEOREM 5. *For $j \neq k \neq l \neq m$, $j, k, l, m = 1, \dots, n_i$, let the responses y_{ij} , y_{ik} , y_{il} , and y_{im} be generated following the four-dimensional joint pdf $L_{i,jklm}$ obtained from (2.4). Then for positive integers r, s, t , and u ,*

$$\begin{aligned} E(Y_{ij}^r Y_{ik}^s Y_{il}^t Y_{im}^u) &= H_{ijklm, (r+s+t+u)}^{(r,s,t,u)} + \frac{\sigma^2}{2} D_{ijklm, (r+s+t+u+2)}^{(r,s,t,u)} \\ &\quad + \frac{\delta_2(\sigma^2)}{6} D_{ijklm, (r+s+t+u+3)}^{(r,s,t,u)} \\ &\quad + \frac{\delta_4(\sigma^2)}{24} D_{ijklm, (r+s+t+u+4)}^{(r,s,t,u)}, \end{aligned} \quad (4.14)$$

where $D_{ijklm, (r+s+t+u+v)}^{(r,s,t,u)}$ are expressed in terms of $H_{ijklm, (r+s+t+u+v)}^{(r,s,t,u)}$ as in Theorem 4, with

$$\begin{aligned} H_{ijklm, (r+s+t+u+v)}^{(r,s,t,u)} &= \int y_{ij}^r y_{ik}^s y_{il}^t y_{im}^u A_{i,jklm}^v f_{ij}(\cdot) \\ &\quad f_{ik}(\cdot) f_{il}(\cdot) f_{im}(\cdot) dy_{ij} dy_{ik} dy_{il} dy_{im}, \end{aligned} \quad (4.15)$$

where for $v = 0, 1, 2, 3$, and 4 , the H functions are provided in the Appendix D.

As mentioned earlier, by putting $r = s = t = u = 1$ in Theorem 5, we obtain the $E(Y_{ij}Y_{ik}Y_{il}Y_{im})$, for $j \neq k \neq l \neq m$, $j, k, l, m = 1, \dots, n_i$. It then follows that

$$\text{cov}(Y_{ij}Y_{ik}, Y_{il}Y_{im}) = E(Y_{ij}Y_{ik}Y_{il}Y_{im}) - E(Y_{ij}Y_{ik})E(Y_{il}Y_{im}), \quad (4.16)$$

where $E(Y_{ij}Y_{ik})$ and $E(Y_{il}Y_{im})$ are given in (3.9) in Theorem 2. This concludes the computation of the covariance matrix

$$M_i^* = \begin{bmatrix} M_{i,2} & M_{i,3}^{(1)} & M_{i,3}^{(2)} \\ & M_{i,4}^{(1)} & M_{i,4}^{(2)} \\ & & M_{i,4}^{(3)} \end{bmatrix},$$

which has been used to construct the estimating equation (4.2) for σ^2 , the variance component of the random effects.

5. TESTS FOR VARIANCE COMPONENT OF THE RANDOM EFFECTS

As the mixed effects model reduces to a fixed effects model when $\sigma^2 = 0$, it may be of interest to test the null hypothesis $H_0: \sigma^2 = 0$ against the alternative hypothesis $H_1: \sigma^2 > 0$. Among the existing tests, the well-known likelihood ratio test appears to be complicated in testing this hypothesis, as in addition to the estimation of β under H_0 it also requires the joint likelihood estimation of β and σ^2 under H_1 . In the following, we provide two alternative asymptotic tests, namely the Wald-type quasi-likelihood test and the score test. These two tests are asymptotically equivalent to the likelihood ratio test which is asymptotically optimal.

THEOREM 6. *The null hypothesis $H_0: \sigma^2 = 0$ against $H_1: \sigma^2 > 0$ may be tested by using the Wald-type test statistic $W_1(\hat{\sigma}_{\text{MQL}}^2)$ given by*

$$W_1(\hat{\sigma}_{\text{MQL}}^2) = [\hat{\sigma}_{\text{MQL}}^2 \{ \hat{V}_{\sigma^2} / I \}^{-1} \hat{\sigma}_{\text{MQL}}^2] \quad (5.1)$$

which under $H_0: \sigma^2 = 0$ has an asymptotic (as $I \rightarrow \infty$) χ^2 distribution with 1 degree of freedom.

Recall from Section 4 that $\hat{\sigma}_{\text{MQL}}^2$ in (5.1) is a consistent estimate of σ^2 and its asymptotic variance is consistently estimated by \hat{V}_{σ^2}/I , where \hat{V}_{σ^2} is given by (4.4). Now the theorem follows from the fact that $W_1(\hat{\sigma}_{\text{MQL}}^2)$ is indeed a Wald-type quasi-likelihood test statistic. This is because $\hat{\sigma}_{\text{MQL}}^2$ is obtained from the quasi-likelihood estimating equation (4.2) and the inverse of its variance, that is, $IV_{\sigma^2}^{-1}$, is equivalent to the well-known Fisher-information matrix derived from the quasi-likelihood of the data instead of the true likelihood.

Note that the computation of $W_1(\hat{\sigma}_{\text{MQL}}^2)$ by (5.1) is also not quite simple as it requires lengthy but straightforward algebras. This is, however, easier to compute as compared to the likelihood ratio test statistic which we have not discussed in this paper. Alternately, one may like to use the score test which is asymptotically equivalent to the Wald and the likelihood ratio tests, and may be easier to compute as it requires only the estimation of the nuisance parameter β under the $H_0: \sigma^2 = 0$. We provide this test in Theorem 7 below.

THEOREM 7. *Let $l = \log \{ \prod_{i=1}^I L_i \}$ denote the log-likelihood function of β and σ^2 , L_i being given in (2.4). Also, let $T(\sigma^2 = 0, \hat{\beta}_{\text{MLE}}^{(0)}) = \partial l / \partial \sigma^2$ be the score function for σ^2 obtained under the $H_0: \sigma^2 = 0$ and evaluated at*

$\beta = \hat{\beta}_{\text{MLE}}^{(0)}$, where $\hat{\beta}_{\text{MLE}}^{(0)}$ is the maximum likelihood estimate of β under the $H_0: \sigma^2 = 0$. Then Rao's (1948) efficient score test statistic is given by

$$W_2(\sigma^2 = 0, \hat{\beta}_{\text{MLE}}^{(0)}) = T^2 / \hat{M}_{11}, \quad (5.2)$$

where \hat{M}_{11} is obtained from $M_{11} = -E(\partial^2 l / \partial \sigma^4)$ as in (5.7) below, evaluating under $H_0: \sigma^2 = 0$, and by replacing β with $\hat{\beta}_{\text{MLE}}^{(0)}$. Under the H_0 , the test statistic $W_2(\sigma^2 = 0, \hat{\beta}_{\text{MLE}}^{(0)})$ has an asymptotic (as $I \rightarrow \infty$) χ^2 distribution with 1 degree of freedom.

The theorem follows from the fact that, for known β , M_{11} is the variance of T .

We now provide the formulas necessary to compute the test statistic $W_2(\cdot)$ in (5.2). First, under the $H_0: \sigma^2 = 0$, it follows from (2.4) that the likelihood estimating equation for β is given by

$$\sum_{i=1}^I \sum_{j=1}^{n_i} [y_{ij} - a'_{ij}] x_{ij} = 0. \quad (5.3)$$

The solution of (5.3), that is $\hat{\beta}_{\text{MLE}}^{(0)}$, may be obtained by the customary Newton-Raphson method. Given the value $\hat{\beta}_{\text{MLE}}^{(0)}(t)$ at the t th iteration, $\hat{\beta}_{\text{MLE}}^{(0)}(t+1)$ is obtained as

$$\hat{\beta}_{\text{MLE}}^{(0)}(t+1) = \hat{\beta}_{\text{MLE}}^{(0)}(t) + \frac{1}{\phi} \left[\sum_{i=1}^I \sum_{j=1}^{n_i} a''_{ij} x_{ij} x_{ij}^T \right]_t^{-1} \left[\sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - a'_{ij}) x_{ij} \right]_t, \quad (5.4)$$

where $[\cdot]_t$ denotes that the expression within brackets is evaluated at $\hat{\beta}_{\text{MLE}}^{(0)}(t)$. The iterative algorithm (5.4) would converge in a manner similar to that of (3.16).

Next we provide the formulas for $T(\sigma^2 = 0, \hat{\beta}_{\text{MLE}}^{(0)})$ and \hat{M}_{11} , which are functions of $\hat{\beta}_{\text{MLE}}^{(0)}$. It is clear from (2.4) that the score equation $T(\cdot)$ for σ^2 evaluated at $\sigma^2 = 0$ and $\beta = \hat{\beta}_{\text{MLE}}^{(0)}$ is given by

$$\begin{aligned} T(\sigma^2 = 0, \hat{\beta}_{\text{MLE}}^{(0)}) &= \{ \partial l / \partial \sigma^2 \}_{|\sigma^2=0, \beta=\hat{\beta}_{\text{MLE}}^{(0)}} \\ &= \frac{\phi}{2} \sum_{i=1}^I \{ \phi A_i^2 - B_i \}, \end{aligned} \quad (5.5)$$

where $A_i = \sum_{j=1}^{n_i} b_{ij} (y_{ij} - a'_{ij})$ and $B_i = \sum_{j=1}^{n_i} b_{ij} a''_{ij}$ are as in (2.4). Further, for given β , it follows that the second derivative of the log likelihood function l with respect to σ^2 is given by

$$\frac{\partial^2 l}{\partial \sigma^4} = \phi \sum_{i=1}^I [\{ (1 + w_i) t_{i2}^* - \phi t_{i1}^{*2} \} / \{ 1 + w_i \}^2], \quad (5.6)$$

where, following (2.4),

$$w_i = \phi \left\{ \frac{\sigma^2}{2} t_{i1} + \frac{\delta_3(\sigma^2)}{6} t_{i2} + \frac{\delta_4(\sigma^2)}{24} t_{i3} \right\},$$

$$t_{i1}^* = \frac{t_{i1}}{2} + \frac{\delta'_3(\sigma^2)}{6} t_{i2} + \frac{\delta'_4(\sigma^2)}{24} t_{i3}, \quad \text{and} \quad t_{i2}^* = \frac{\delta''_3(\sigma^2)}{6} t_{i2} + \frac{\delta''_4(\sigma^2)}{24} t_{i3},$$

with

$$t_{i1} = \phi A_i^2 - B_i, \quad t_{i2} = \phi^2 A_i^3 - 3\phi A_i B_i - C_i,$$

and

$$t_{i3} = \phi^3 A_i^4 - 6\phi^2 A_i^2 B_i - 4\phi A_i C_i + 3\phi B_i^2 - D_i.$$

Now, by using the approximate mean of the ratio of two variables U and V as $E(U/V) \approx E(U)/E(V)$, we compute, after some lengthy algebras, the expectation of $\partial^2 l / \partial \sigma^4$ (5.6), evaluated at $\sigma^2 = 0$, as

$$\begin{aligned} M_{11} &= -E(\partial^2 l / \partial \sigma^4) \\ &\approx \phi \sum_{i=1}^I \left[\phi \left\{ \sum_{s=1}^8 w_{si} m_{si}^* + C_{0i}^* \right\} \right. \\ &\quad \left. - \frac{c_{4,1}}{12} \{ \phi^3 m_{4i}^* - 6\phi^3 m_{2i}^* B_i + 3\phi B_i^2 - D_i \} \right], \end{aligned} \quad (5.7)$$

where $m_{1i}^* = 0$ and for $s = 2, \dots, 8$,

$$m_{si}^* = \sum_{j=1}^{n_i} b_{ij}^{s/2} m_{ij,s},$$

with $m_{ij,s}$ ($s = 2, \dots, 8$) as in Lemma 1. The normalizing constant C_{0i}^* and the weights w_{si} in (5.7), are given by

$$\begin{aligned} C_{0i}^* &= (1/4) B_i^2 + \{c_{3,2}^2/36\} C_i^2 + \{c_{3,2}/6\} B_i C_i + \{c_{4,3}^2/576\} (3\phi B_i^2 - D_i)^2 \\ &\quad + \{c_{4,3}/24\} (3\phi B_i^2 - D_i) B_i - \{c_{3,2} c_{4,3}/72\} (3\phi B_i^2 - D_i) C_i, \\ w_{2i} &= -(1/2) \phi B_i + (c_{3,2}^2/4) \phi^2 B_i^2 + (c_{4,3}^2/36) \phi^2 C_i^2 \\ &\quad - (c_{4,3}^2/48) (3\phi B_i^2 - D_i) \phi^2 B_i \\ &\quad - (c_{3,2}/6) \phi C_i + (3c_{4,3}/8) \phi^2 B_i^2 - (c_{4,3}/24) \phi D_i + (c_{3,2}/4) \phi^2 B_i C_i, \end{aligned}$$

$$\begin{aligned}
w_{3i} &= -(c_{3,2}^2/18) \phi^2 C_i + (c_{4,3}^2/24) \phi^3 B_i C_i - c_{3,2} \phi^2 B_i \\
&\quad - (c_{4,3}/6) \phi^2 C_i + (21c_{3,2}c_{4,3}/72) \phi^3 B_i^2 - (c_{3,2}c_{4,3}/72) \phi^2 D_i, \\
w_{4i} &= (1/4) \phi^2 - (c_{3,2}^2/6) \phi^3 B_i + (42c_{4,3}^2/576) \phi^4 B_i^2 - (c_{4,3}^2/288) \phi^3 D_i \\
&\quad - (7c_{4,3}/24) \phi^3 B_i - (5c_{3,2}c_{4,3}/72) \phi^3 C_i, \\
w_{5i} &= -(c_{4,3}^2/144) \phi^4 C_i + (c_{3,2}/6) \phi^3 - (c_{3,2}c_{4,3}/8) \phi^4 B_i, \\
w_{6i} &= (c_{3,2}^2/36) \phi^4 - (c_{4,3}^2/48) \phi^5 B_i + (c_{4,3}/24) \phi^4,
\end{aligned}$$

and

$$w_{7i} = (c_{3,2}c_{4,3}/72) \phi^5, \text{ and } w_{8i} = (c_{4,3}^2/576) \phi^6,$$

where, for $r = 1, \dots, 4$ and $u = 1, \dots, r$, $c_{r,u}$ are given by (2.2).

6. TESTING THE HOMOGENEITY OF VARIANCES

In this section, we consider the testing for the homogeneity of the variance components of the random effects of several conceptual groups, where the clustered observations in each group follow the GLMMs with the same regression parameters but with possibly different variance components of the random effects. Lin (1997) considers a similar but different testing problems. More specifically, Lin (1997) considers only one conceptual group with a large number of clusters, where clustered observations follow certain factorial designs with several factors, and tests the homogeneity of the variance components due to these factors.

Let there be G independent conceptual groups. In practice, G is usually small, say 2 or 3. Let there be I_g clusters in the g th ($g = 1, \dots, G$) conceptual group and n_{gi} the number of observations in the i th ($i = 1, \dots, I_g$) cluster of the g th ($g = 1, \dots, G$) group. Further, let σ_g^2 be the variance of the random effects in the g th group. We then wish to test the hypothesis

$$H_0: \sigma_1^2 = \dots = \sigma_g^2 = \dots = \sigma_G^2 = 0, \quad (6.1)$$

where it is assumed that n_{gi} observations of the i th cluster in the g th group are generated following the GLMMs discussed in Section 2. Following (2.4), the approximate likelihood function for $\beta, \sigma_1^2, \dots, \sigma_G^2$ may be written as

$$\begin{aligned}
L^+(\beta, \sigma_1^2, \dots, \sigma_G^2) &= \prod_{g=1}^G \prod_{i=1}^{I_g} [\{ f(y_{ij} | \theta_{ij}) \} [1 + \phi \{ (\sigma_g^2/2) t_{gi1} \\
&\quad + \{ \delta_3(\sigma_g^2)/6 \} t_{gi2} + \{ \delta_4(\sigma_g^2)/24 \} t_{gi3} \}]], \quad (6.2)
\end{aligned}$$

where t_{gi1} 's are defined as in (5.6). For example,

$$t_{gi3} = \phi^3 A_{gi}^4 - 6\phi^2 A_{gi}^2 B_{gi} - 4\phi A_{gi} C_{gi} + 3\phi B_{gi}^2 - D_{gi},$$

where

$$\begin{aligned} A_{gi} &= \sum_{j=1}^{n_{gi}} b_{gij}^{1/2} (y_{gij} - a'_{gij}), & B_{gi} &= \sum_{j=1}^{n_{gi}} b_{gij} a''_{gij}, \\ C_{gi} &= \sum_{j=1}^{n_{gi}} b_{gij}^{3/2} a'''_{gij}, & \text{and} & \quad D_{gi} = \sum_{j=1}^{n_{gi}} b_{gij}^2 a^{IV}_{gij}, \end{aligned}$$

y_{gij} being the j th observation of the i th cluster in the g th ($g = 1, \dots, G$) group.

The null hypothesis, $H_0: \sigma_1^2 = \dots = \sigma_g^2 = \dots = \sigma_G^2 = 0$ in (6.1) may now be tested by using the score test statistic

$$W_{2G}(\sigma_1^2 = 0, \dots, \sigma_G^2 = 0, \hat{\beta}_{MLE}^{(0)}) = \sum_{g=1}^G T_g^2 / \hat{M}_{11g}, \quad (6.3)$$

where, similar to (5.5), $T_g = (\phi/2) \sum_{i=1}^{I_g} \{\phi A_{gi}^2 - B_{gi}\}$, which is evaluated at $\hat{\beta}_{MLE}^{(0)}$, the maximum likelihood estimate of β under $H_0: \sigma_1^2 = \dots = \sigma_G^2 = 0$. Here $\hat{\beta}_{MLE}^{(0)}$ is the solution of the likelihood estimating equation

$$\sum_{g=1}^G \sum_{i=1}^{I_g} \sum_{j=1}^{n_{gi}} [y_{gij} - a'_{gij}] x_{gij} = 0. \quad (6.4)$$

In (6.3), M_{11g} is computed following (5.7). That is,

$$\begin{aligned} M_{11g} &\approx \phi \sum_{i=1}^{I_g} \left[\phi \left\{ \sum_{s=1}^8 w_{gsi} m_{gsi}^* + C_{goi}^* \right\} \right. \\ &\quad \left. - \frac{c_{4,1}}{12} \{ \phi^3 m_{g4i}^* - 6\phi^3 m_{g2i}^* B_{gi} + 3\phi B_{gi}^2 - D_{gi} \} \right], \end{aligned} \quad (6.5)$$

where m_{gsi}^* are obtained from m_{si}^* by replacing b_{ig} and $m_{ij,s}$ with b_{gij} and $m_{gij,s}$ respectively. Similarly, C_{goi}^* and w_{gsi} are obtained from C_{oi}^* and w_{si} respectively, by replacing B_i , C_i , and D_i with B_{gi} , C_{gi} and D_{gi} respectively.

Under $H_0: \sigma_1^2 = \dots = \sigma_g^2 = \dots = \sigma_G^2 = 0$, the test statistic $W_{2G}(\cdot)$ in (6.3) has an asymptotic (as $I_g \rightarrow \infty$, for each $g = 1, \dots, G$) χ^2 distribution with G degrees of freedom. Note that the hypothesis $H_0: \sigma_1^2 = \dots = \sigma_g^2 = \dots = \sigma_G^2 = \tilde{\sigma}^2$, say, may be of interest for testing as well, which can be tested using a similar approach, after replacing $\sigma_g^2 (g = 1, \dots, G)$ with an estimate of $\tilde{\sigma}^2$. Consequently, in this case, the test statistic $W_{2G}(\cdot)$ in (6.3) will have a χ^2 distribution with $G - 1$ degrees of freedom.

7. CONCLUDING REMARKS

Breslow and Clayton (1993) have used the so-called penalized quasi-likelihood (PQL) approach to estimate the regression coefficients and variance component σ^2 of the random effects in generalized linear mixed models (GLMMs). This approach generally produces biased estimates for the regression intercept and the variance component parameters. For small values of the variance component, Breslow and Lin (1995) proposed a bias correction to such PQL estimators. Recently, Sutradhar and Qu (1998) have proposed a small variance component based likelihood approximation (LA) to estimate the parameters of a Poisson mixed model and compared the performance of the LA estimators with that of the PQL estimators of Breslow and Clayton as well as the so-called Stein-type estimating function (SEF) based estimators of Waclawiw and Liang (1993). It has been shown in Sutradhar and Qu (1998) that the LA approach yields consistent estimates for small σ^2 , the SEF approach never produces consistent estimates for σ^2 , and the PQL approach may or may not yield consistent estimates, depending on the cluster size n_i and the design matrix (x_{ij}) . Note, however, that there does not appear to be any immediate generalization of Sutradhar and Qu's approach for the Poisson mixed models to the generalized linear mixed models.

An alternative method of estimation of the parameters of the GLMMs is the so-called marginal quasi-likelihood (MQL) method. This method has been used by several authors including Breslow and Clayton (1993), Prentice and Zhao (1991), Liang *et al.* (1992), and Sutradhar and Rao (1996). The MQL approach is, however, hampered by the requirement that the marginal moments of the responses up to order four need to be known. But these moments are not easy to compute under the generalized linear mixed models. Following Zeger *et al.* (1988), Breslow and Clayton (1993) use the MQL approach, based on certain approximations to the mean and the covariance matrix. But these approximations as well as the unavailability of the third and the fourth order moments have adverse effects on the efficiency of the estimates. As a remedy to this problem, this paper first developed a small σ^2 based joint likelihood (cf. Lin, 1997) of the responses in a cluster. Next, instead of maximizing the likelihood, this paper has exploited the likelihood function to obtain the moments of the responses up to order four. These moments are then used to develop the MQL estimating equations for the regression as well as the variance component of the random effects. This four-moment-based MQL approach provides consistent as well as more efficient estimates for the regression and the overdispersion parameters as compared to the MQL estimators computed based on the approximate mean and the working second, third, fourth order moments of the responses. The numerical computations with regard

to the efficiency gain shown in Section 3 fully support this conclusion for the regression estimates computed based on the improved second moments.

The paper also develops the score tests for testing the homogeneity of variances of the random effects of several conceptual groups following the GLMMs. The test statistic is simple to compute and it has an asymptotically valid χ^2 distribution with G degrees of freedom, where G is the number of independent groups with possibly unequal variances for the random effects.

APPENDIX A

Derivative of the D_i Matrix

In (3.15), $D_i = \partial\{M'_{i,1}\}/\partial\beta$ is the $n_i \times p$ first derivative matrix of $M'_{i,1}$ with respect to β . From (3.13),

$$M'_{i,1} = [M'_{i1,1}, \dots, M'_{ij,1}, \dots, M'_{in_i,1}]^T,$$

with $E(Y_{ij}) = M'_{ij,1}$ given by (3.10). Since $\theta_{ij} = x_{ij}^T \beta$ and a'_{ij}, \dots, a^V_{ij} are, respectively, the first five order derivatives of a_{ij} with respect to θ_{ij} , it then follows from (3.10) that

$$\begin{aligned} \partial M'_{ij,1} / \partial \beta &= \left[a''_{ij} + \frac{\sigma^2}{2} b_{ij} a_{ij}^{\text{IV}} + \frac{\delta^3(\sigma^2)}{6} b_{ij}^{3/2} a_{ij}^V + \frac{\delta_4(\sigma^2)}{24} b_{ij}^2 a_{ij}^{\text{VI}} \right] x_{ij}^T, \\ &= z_{ij} x_{ij}^T, \end{aligned}$$

so that

$$\begin{aligned} D_i &= \partial M'_{i,1} / \partial \beta = \begin{bmatrix} z_{i1} x_{i1}^T \\ \vdots \\ z_{in_i} x_{in_i}^T \end{bmatrix} = \begin{pmatrix} z_{in_i} & 0 & \cdots & 0 \\ 0 & z_{i2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & z_{in_i} \end{pmatrix} X_i \\ &= Z_i X_i, \end{aligned}$$

say, where $Z_i = \text{diag}[z_{i1}, \dots, z_{in_i}]$ and $X_i = [x_{i1}, \dots, x_{ij}, \dots, x_{in_i}]^T$.

APPENDIX B

Derivative of the D_i^ Matrix*

Recall from (4.2) that

$$D_i^* = \partial[M'_{i,1}^T, m_{i1,2}^{*T}, m_{i2,2}^{*T}]^T / \partial \sigma^2,$$

where $M'_{i,1}{}^T$, $m_{i1,2}^{*\top}$, and $m_{i2,2}^{*\top}$ are, respectively, the vectors of order $1 \times n_i$, $1 \times n_i$, and $1 \times \{n_i(n_i - 1)/2\}$.

Now it follows from (3.10) that

$$\frac{\partial M'_{ij,1}}{\partial \sigma^2} = (1/2) b_{ij} a''_{ij} + \frac{\delta'_3(\sigma^2)}{6} b_{ij}^{3/2} a_{ij}^{\text{IV}} + \frac{\delta'_4(\sigma^2)}{24} b_{ij}^2 a_{ij}^{\text{V}},$$

for all $j = 1, \dots, n_i$. Similarly, as the j th element of the $1 \times n_i$ vector $m_{i1,2}^{*\top}$ is given by $M'_{ij,2} = E(Y_{ij}^2)$, it follows from (3.5) and (3.11) that

$$\frac{\partial M'_{ij,2}}{\partial \sigma^2} = \frac{\partial M_{ij,2}}{\partial \sigma^2} + \frac{\partial}{\partial \sigma^2} [M'_{ij,1}]^2,$$

where,

$$\frac{\partial}{\partial \sigma^2} [M'_{ij,1}]^2 = 2M'_{ij,1} \left(\frac{\partial M'_{ij,1}}{\partial \sigma^2} \right)$$

and, by (3.11),

$$\begin{aligned} \frac{\partial M_{ij,2}}{\partial \sigma^2} &= b_{ij} \{ a'_{ij} a'''_{ij} + (a''_{ij})^2 + (1/2\phi) a_{ij}^{\text{IV}} \} \\ &\quad + \frac{\delta'_3(\sigma^2)}{3} b_{ij}^{3/2} \{ a'_{ij} a_{ij}^{\text{IV}} + 3a''_{ij} a'''_{ij} + (1/2\phi) a_{ij}^{\text{V}} \} \\ &\quad + \frac{\delta'_4(\sigma^2)}{12} b_{ij}^2 \{ a'_{ij} a_{ij}^{\text{V}} + 4a''_{ij} a_{ij}^{\text{IV}} + 3(a'''_{ij})^2 + (1/2\phi) a_{ij}^{\text{VI}} \}. \end{aligned}$$

Next, for $j < k$, $j, k = 1, \dots, n_i$, the (j, k) th element of the $1 \times \{n_i(n_i - 1)/2\}$ vector $m_{i2,2}^{*\top}$ is given by $M'_{ijk,2} = E(Y_{ij} Y_{ik})$. It then follows from (3.9) and (3.12) that

$$\frac{\partial M'_{ijk,2}}{\partial \sigma^2} = \frac{\partial M_{ijk,2}}{\partial \sigma^2} + \frac{\partial}{\partial \sigma^2} [M'_{ij,1} M'_{ik,1}],$$

where,

$$\frac{\partial}{\partial \sigma^2} [M'_{ij,1} M'_{ik,1}] = M'_{ij,1} \frac{\partial}{\partial \sigma^2} M'_{ik,1} + M'_{ik,1} \frac{\partial}{\partial \sigma^2} M'_{ij,1},$$

and by (3.12),

$$\begin{aligned}
\frac{\partial M_{ijk,2}}{\partial \sigma^2} = & (1/2) \{ b_{ij} a_{ij}''' a'_{ik} + 2b_{ij}^{1/2} b_{ik}^{1/2} a''_{ij} a''_{ik} + b_{ik} a'_{ij} a_{ik}''' \} \\
& + \frac{\delta'_3(\sigma^2)}{6} \{ b_{ij}^{3/2} a_{ij}^{\text{IV}} a'_{ik} + 3b_{ij} b_{ik}^{1/2} a_{ij}''' a''_{ik} \\
& + 3b_{ij}^{1/2} b_{ik} a''_{ij} a_{ik}''' + b_{ij}^{3/2} a'_{ij} a_{ik}^{\text{IV}} \} \\
& + \frac{\delta'_4(\sigma^2)}{24} \{ b_{ij}^2 a_{ij}^{\text{V}} a'_{ik} + 4b_{ij}^{3/2} b_{ik}^{1/2} a_{ij}^{\text{IV}} a''_{ik} + 6b_{ij} b_{ik} a_{ij}''' a_{ik}''' \\
& + 4b_{ij}^{1/2} b_{ik}^{3/2} a_{ij}'' a_{ik}^{\text{IV}} + b_{ik}^2 a'_{ij} a_{ik}^{\text{V}} \}. \tag{10.2}
\end{aligned}$$

Finally, by combining the above results, one obtains the D_i^* vector.

APPENDIX C

Formulas for H Functions in Theorem 4

By direct integration, after some lengthy algebras, it follows from (4.7) that

$$\begin{aligned}
H_{ijkl, (r+s+t)}^{(r+s+t)} &= h_{ij, (r)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t)}^{(t)} \\
H_{ijkl, (r+s+t+1)}^{(r, s, t)} &= h_{ij, (r+1)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (s)}^{(t)} \\
&\quad + h_{ij, (r)}^{(r)} h_{ik, (s+1)}^{(t)} h_{il, (t)}^{(t)} + h_{ij, (r)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+1)}^{(t)} \\
H_{ijkl, (r+s+t+2)}^{(r, s, t)} &= [h_{ij, (r+2)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t)}^{(t)} + h_{ij, (r)}^{(r)} h_{ik, (s+2)}^{(s)} h_{il, (t)}^{(t)} \\
&\quad + h_{ij, (r)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+2)}^{(t)}] \\
&\quad + 2[h_{ij, (r+1)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t)}^{(t)} + h_{ij, (r+1)}^{(r)} h_{ik, s}^{(s)} h_{il, (t+1)}^{(t)} \\
&\quad + h_{ij, (r)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t+1)}^{(t)}], \\
H_{ijkl, (r+s+t+3)}^{(r, s, t)} &= [h_{ij, (r+3)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t)}^{(t)} + h_{ij, (r)}^{(r)} h_{ik, (s+3)}^{(s)} h_{il, (t)}^{(t)} \\
&\quad + h_{ij, (r)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+3)}^{(t)}] \\
&\quad + 3[h_{ij, (r+2)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t)}^{(t)} + h_{ij, (r+2)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+1)}^{(t)} \\
&\quad + h_{ij, (r+1)}^{(r)} h_{ik, (s+2)}^{(s)} h_{il, (t)}^{(t)} \\
&\quad + h_{ij, (r+1)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+2)}^{(t)} + h_{ij, (r)}^{(r)} h_{ik, (s+2)}^{(s)} h_{il, (t+1)}^{(t)} \\
&\quad + h_{ij, (r)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t+2)}^{(t)}] \\
&\quad + 6[h_{ij, (r+2)}^{(r)} h_{ik, (s+2)}^{(s)} h_{il, (t+2)}^{(t)}],
\end{aligned}$$

$$\begin{aligned}
H_{ijkl, (r+s+t+4)}^{(r, s, t)} &= [h_{ij, (r+4)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t)}^{(t)} \\
&\quad + h_{ij, (r)}^{(r)} h_{ik, (s+4)}^{(s)} h_{il, (t)}^{(t)} + h_{ij, (r)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+4)}^{(t)}] \\
&\quad + 4[h_{ij, (r+3)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t)}^{(t)} + h_{ij, (r+3)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+1)}^{(t)} \\
&\quad + h_{ij, (r+1)}^{(r)} h_{ik, (s+3)}^{(s)} h_{il, (t)}^{(t)} \\
&\quad + h_{ij, (r+1)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+3)}^{(t)} + h_{ij, (r)}^{(r)} h_{ik, (s+3)}^{(s)} h_{il, (t+1)}^{(t)} \\
&\quad + h_{ij, (r)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t+3)}^{(t)}] \\
&\quad + 12[h_{ij, (r+2)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t+1)}^{(t)} + h_{ij, (r+1)}^{(r)} h_{ik, (s+2)}^{(s)} h_{il, (t+1)}^{(t)} \\
&\quad + h_{ij, (r+1)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t+2)}^{(t)}].
\end{aligned}$$

APPENDIX D

Formulas for H Functions in Theorem 5

By integrations similar to these in (4.7), it now follows from (4.15) that

$$\begin{aligned}
H_{ijklm, (r+s+t+u)}^{(r, s, t, u)} &= h_{ij, (r)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u)}^{(u)}, \\
H_{ijklm, (r+s+t+u+1)}^{(r, s, t, u)} &= h_{ij, (r+1)}^{(r)} h_{ij, (s)}^{(s)} h_{ik, (t)}^{(t)} h_{il, (u)}^{(u)} \\
&\quad + h_{ij, (r)}^{(r)} h_{ij, (s+1)}^{(s)} h_{ik, (t)}^{(t)} h_{il, (u)}^{(u)} \\
&\quad + h_{ij, (r)}^{(r)} h_{ij, (s)}^{(s)} h_{ij, (t+1)}^{(t)} h_{il, (u)}^{(u)} \\
&\quad + h_{ij, (r)}^{(r)} h_{ij, (s)}^{(s)} h_{ij, (t)}^{(t)} h_{il, (u+1)}^{(u)}, \\
H_{ijklm, (r+s+t+u+2)}^{(r, s, t, u)} &= h_{ij, (r+2)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u)}^{(u)} \\
&\quad + h_{ij, (r)}^{(r)} h_{ik, (s+2)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u)}^{(u)} \\
&\quad + h_{ij, (r)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+2)}^{(t)} h_{im, (u)}^{(u)} \\
&\quad + h_{ij, (r)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u+2)}^{(u)} \\
&\quad + 2[h_{ij, (r+1)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u)}^{(u)} \\
&\quad + h_{ij, (r+1)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+1)}^{(t)} h_{im, (u)}^{(u)} \\
&\quad + h_{ij, (r+1)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u+1)}^{(u)} \\
&\quad + h_{ij, (r)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t+1)}^{(t)} h_{im, (u)}^{(u)} \\
&\quad + h_{ij, (r)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u+1)}^{(u)} \\
&\quad + h_{ij, (r)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+1)}^{(t)} h_{im, (u+1)}^{(u)}],
\end{aligned}$$

$$\begin{aligned}
H_{ijklm, (r+s+t+u+3)}^{(r,s,t,u)} = & h_{ij, (r+3)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u)}^{(u)} \\
& + h_{ij, (r)}^{(r)} h_{ik, (s+3)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u)}^{(u)} \\
& + h_{ij, (r)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+3)}^{(t)} h_{im, (u)}^{(u)} \\
& + h_{ij, (r)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u+3)}^{(u)} \\
& + 3[h_{ij, (r+2)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u)}^{(u)} \\
& + h_{ij, (r+2)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+1)}^{(t)} h_{im, (u)}^{(u)} \\
& + h_{ij, (r+2)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u+1)}^{(u)} \\
& + h_{ij, (r+1)}^{(r)} h_{ik, (s+2)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u)}^{(u)} \\
& + h_{ij, (r+1)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+2)}^{(t)} h_{im, (u)}^{(u)} \\
& + h_{ij, (r+1)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u+2)}^{(u)}] \\
& + 3[h_{ij, (r)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+2)}^{(t)} h_{im, (u+1)}^{(u)} \\
& + h_{ij, (r)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+1)}^{(t)} h_{im, (u+2)}^{(u)} \\
& + h_{ij, (r)}^{(r)} h_{ik, (s+2)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u+1)}^{(u)} \\
& + h_{ij, (r)}^{(r)} h_{ik, (s+2)}^{(s)} h_{il, (t+1)}^{(t)} h_{im, (u)}^{(u)} \\
& + h_{ij, (r)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u+2)}^{(u)} \\
& + h_{ij, (r)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t+2)}^{(t)} h_{im, (u)}^{(u)}] \\
& + 6[h_{ij, (r+1)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t+1)}^{(t)} h_{im, (u)}^{(u)} \\
& + h_{ij, (r+1)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u+1)}^{(u)} \\
& + h_{ij, (r+1)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+1)}^{(t)} h_{im, (u+1)}^{(u)} \\
& + h_{ij, (r)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t+1)}^{(t)} h_{im, (u+1)}^{(u)}],
\end{aligned}$$

and

$$\begin{aligned}
H_{ijklm, (r+s+t+u+4)}^{(r,s,t,u)} = & h_{ij, (r+4)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u)}^{(u)} \\
& + h_{ij, (r)}^{(r)} h_{ik, (s+4)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u)}^{(u)} \\
& + h_{ij, (r)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+4)}^{(t)} h_{im, (u)}^{(u)} \\
& + h_{ij, (r)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u+4)}^{(u)} \\
& + 4[h_{ij, (r+3)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u)}^{(u)} \\
& + h_{ij, (r+3)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+1)}^{(t)} h_{im, (u)}^{(u)} \\
& + h_{ij, (r+3)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u+1)}^{(u)}
\end{aligned}$$

$$\begin{aligned}
& + h_{ij, (r+1)}^{(r)} h_{ik, (s+3)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u)}^{(u)} \\
& + h_{ij, (r+1)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+3)}^{(t)} h_{im, (u)}^{(u)} \\
& + h_{ij, (r+1)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u+3)}^{(u)}] \\
& + 4[h_{ij, (r)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+3)}^{(t)} h_{im, (u+1)}^{(u)} \\
& + h_{ij, (r)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+1)}^{(t)} h_{im, (u+3)}^{(u)} \\
& + h_{ij, (r)}^{(r)} h_{ik, (s+3)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u+1)}^{(u)} \\
& + h_{ij, (r)}^{(r)} h_{ik, (s+3)}^{(s)} h_{il, (t+1)}^{(t)} h_{im, (u)}^{(u)} \\
& + h_{ij, (r)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u+3)}^{(u)} \\
& + h_{ij, (r)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t+3)}^{(t)} h_{im, (u)}^{(u)}] \\
& + 6[h_{ij, (r+2)}^{(r)} h_{ik, (s+2)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u)}^{(u)} \\
& + h_{ij, (r+2)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+2)}^{(t)} h_{im, (u)}^{(u)} \\
& + h_{ij, (r+2)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u+2)}^{(u)} \\
& + h_{ij, (r)}^{(r)} h_{ik, (s+2)}^{(s)} h_{il, (t+2)}^{(t)} h_{im, (u)}^{(u)} \\
& + h_{ij, (r)}^{(r)} h_{ik, (s+2)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u+1)}^{(u)} \\
& + h_{ij, (r)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+2)}^{(t)} h_{im, (u+2)}^{(u)}] \\
& + 12[h_{ij, (r+2)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+1)}^{(t)} h_{im, (u+1)}^{(u)} \\
& + h_{ij, (r+2)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u+1)}^{(u)} \\
& + h_{ij, (r+2)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t+1)}^{(t)} h_{im, (u)}^{(u)} \\
& + h_{ij, (r)}^{(r)} h_{ik, (s+2)}^{(s)} h_{il, (t+1)}^{(t)} h_{im, (u+1)}^{(u)} \\
& + h_{ij, (r)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t+2)}^{(t)} h_{im, (u+1)}^{(u)} \\
& + h_{ij, (r)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t+1)}^{(t)} h_{im, (u+2)}^{(u)}] \\
& + 12[h_{ij, (r+1)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t+2)}^{(t)} h_{im, (u)}^{(u)} \\
& + h_{ij, (r+1)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u+2)}^{(u)} \\
& + h_{ij, (r+1)}^{(r)} h_{ik, (s+2)}^{(s)} h_{il, (t+1)}^{(t)} h_{im, (u)}^{(u)} \\
& + h_{ij, (r+1)}^{(r)} h_{ik, (s+2)}^{(s)} h_{il, (t)}^{(t)} h_{im, (u+1)}^{(u)} \\
& + h_{ij, (r+1)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+1)}^{(t)} h_{im, (u+2)}^{(u)} \\
& + h_{ij, (r+1)}^{(r)} h_{ik, (s)}^{(s)} h_{il, (t+2)}^{(t)} h_{im, (u+1)}^{(u)}] \\
& + 24h_{ij, (r+1)}^{(r)} h_{ik, (s+1)}^{(s)} h_{il, (t+1)}^{(t)} h_{im, (u+1)}^{(u)} .
\end{aligned}$$

ACKNOWLEDGMENTS

The authors are grateful to Bhagavan Sri Sathya Sai Baba for providing opportunities to carry out this research at the Sri Sathya Sai Institute of Higher Learning. This work was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada. The authors thank the referee for valuable comments and suggestions leading to the improvement of the paper. We also thank the editor for two additional references.

REFERENCES

1. N. E. Breslow and D. G. Clayton, Approximate inference in generalized linear mixed models, *J. Amer. Statist. Assoc.* **88** (1993), 9–25.
2. N. E. Breslow and X. Lin, Bias correction in generalized linear models with a single component of dispersion, *Biometrika* **82** (1995), 81–92.
3. R. J. Carroll and D. Ruppert, Robust estimation in heteroscedastic linear models, *Ann. Statist.* **10** (1982), 429–441.
4. K. Das and B. C. Sutradhar, Tests for overdispersion in a regression model for time dependent truncated data, *Sankhya A* **58** (1986), 473–490.
5. H. Goldstein and J. Rasbash, Improved approximations for multilevel models with binary responses, *J. Roy. Statist. Soc. Ser. A* **159** (1996), 505–513.
6. K.-Y. Liang and S. L. Zeger, Longitudinal data analysis using generalized linear models, *Biometrika* **73** (1986), 13–22.
7. K.-Y. Liang, S. L. Zeger, and B. Qaqish, Multivariate regression analysis for categorical data, *J. Roy. Statist. Soc. Ser. B* **54** (1992), 3–40.
8. X. Lin, Variance component testing in generalized linear models with random effects, *Biometrika* **84** (1997), 309–326.
9. P. McCullagh, Quasilikelihood functions, *Ann. Statist.* **11** (1983), 59–67.
10. P. McCullagh and J. A. Nelder, “Generalized Linear Models,” Chapman and Hall, London, 1989.
11. R. L. Prentice and L. P. Zhao, Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses, *Biometrics* **47** (1991), 825–839.
12. C. R. Rao, Large sample tests of statistical hypothesis concerning several parameters with applications to problems of estimation, *Proc. Cambridge Philos. Soc.* **44** (1948), 50–57.
13. G. Rodriguez and N. Goldman, An assesment of estimation procedures for multilevel models with binary responses, *J. Roy. Statist. Soc. Ser. A* **158** (1995), 73–89.
14. G. A. F. Seber and C. J. Wild, “Nonlinear Regression,” Wiley, New York, 1989.
15. P. J. Solomon and D. R. Cox, Nonlinear components of variance models, *Biometrika* **79** (1992), 1–11.
16. B. C. Sutradhar and K. Das, Analysing exponential family based messy data, *Comm. Statist. Theory Methods* **24** (1995), 2683–2699.
17. B. C. Sutradhar and Z. Qu, On approximate likelihood inference in Poisson mixed model, *Canad. J. Statist.* **26**, No. 1 (1998), 169–186.
18. B. C. Sutradhar and R. P. Rao, On joint estimation of regression and overdispersion parameters in generalized linear models for longitudinal data, *J. Multivariate Anal.* **56** (1996), 90–119.

19. M. A. Waclawiw and K.-Y. Liang, Prediction of random effects in the generalized linear model, *J. Amer. Statist. Assoc.* **88** (1993), 171–178.
20. R. Wedderburn, Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method, *Biometrika* **61** (1979), 439–447.
21. S. L. Zeger, K.-Y. Liang, and P. S. Albert, Models for longitudinal data: A generalized estimating equation approach, *Biometrics* **44** (1988), 1049–1060.